

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2003, 23:03:01 ; Search time 81 Seconds

(without alignments)
981.904 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038
Sequence: 1 MLPKRMKLLPLVLSQMAIL.....EORNLLDLTLSTWTEQIIR 386

Scoring table: BLAST62
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL.21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.prodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriag:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 2038 | 100.0 | 386 | 4 Q9Y5R3 | Q9Y5R3 homo sapien |
| 2 | 1515.5 | 74.4 | 388 | 11 Q9RL11 | Q9RL11 mus musculu |
| 3 | 1513.5 | 74.3 | 388 | 11 Q9WUE5 | Q9WUE5 mus musculu |
| 4 | 1019 | 50.0 | 395 | 4 Q9G2X3 | Q9G2X3 homo sapien |
| 5 | 1017.5 | 49.9 | 395 | 11 Q9GUP4 | Q9GUP4 mus musculu |
| 6 | 1008 | 49.5 | 411 | 4 Q9GZS9 | Q9GZS9 homo sapien |
| 7 | 627 | 30.8 | 484 | 11 Q9EP78 | Q9EP78 mus musculu |
| 8 | 624 | 30.6 | 484 | 11 Q9NP78 | Q9NP78 mus musculu |
| 9 | 606.5 | 29.8 | 486 | 4 Q7567 | Q7567 homo sapien |
| 10 | 605.5 | 29.7 | 486 | 4 Q9NS84 | Q9NS84 homo sapien |
| 11 | 598.5 | 29.4 | 530 | 11 Q9G276 | Q9G276 mus musculu |
| 12 | 591.5 | 29.0 | 530 | 4 Q9Y6F2 | Q9Y6F2 homo sapien |
| 13 | 587.5 | 28.8 | 484 | 4 Q9UED5 | Q9UED5 homo sapien |
| 14 | 587.5 | 28.8 | 531 | 4 Q9Y4C5 | Q9Y4C5 homo sapien |
| 15 | 559 | 27.4 | 411 | 11 Q9BQC0 | Q9BQC0 mus musculu |
| 16 | 545 | 26.9 | 411 | 4 Q43516 | Q43516 homo sapien |

ALIGNMENTS

| | | | | | |
|----|-------|------|------|-----------|--------------------|
| 17 | 504.5 | 24.8 | 472 | 11 Q88199 | Q88199 mus musculu |
| 18 | 482 | 23.7 | 479 | 4 Q75099 | Q75099 homo sapien |
| 19 | 481.5 | 23.6 | 474 | 11 Q9Q212 | Q9Q212 ratu8 norv |
| 20 | 423.5 | 20.8 | 441 | 13 Q93403 | Q93403 torpedo cal |
| 21 | 264 | 13.0 | 486 | 5 Q95T19 | Q95T19 drosophila |
| 22 | 210 | 10.3 | 363 | 5 Q9VWC4 | Q9VWC4 drosophila |
| 23 | 203 | 10.0 | 114 | 11 Q9D0K5 | Q9D0K5 mus musculu |
| 24 | 186.5 | 9.2 | 183 | 5 Q9VMC2 | Q9VMC2 mus musculu |
| 25 | 175 | 8.6 | 119 | 6 Q95T18 | Q95T18 oryctolagus |
| 26 | 168 | 8.2 | 307 | 16 Q92VCA | Q92VCA tritobium m |
| 27 | 113.5 | 5.7 | 303 | 16 Q93UB6 | Q93UB6 streptomyce |
| 28 | 111 | 5.4 | 655 | 6 Q9B6X0 | Q9B6X0 macaca fasc |
| 29 | 107 | 5.3 | 1586 | 4 Q9P2D3 | Q9P2D3 homo sapien |
| 30 | 100.5 | 4.9 | 747 | 4 Q96OU2 | Q96OU2 homo sapien |
| 31 | 100.5 | 4.9 | 775 | 4 Q9C011 | Q9C011 homo sapien |
| 32 | 100 | 4.9 | 593 | 10 Q8VZE0 | Q8VZE0 arabidopsis |
| 33 | 96.5 | 4.7 | 743 | 4 Q9H0K2 | Q9H0K2 homo sapien |
| 34 | 96 | 4.7 | 2454 | 3 Q9VVP2 | Q9VVP2 emericella |
| 35 | 96 | 4.7 | 2454 | 3 Q9UV56 | Q9UV56 emericella |
| 36 | 92.5 | 4.5 | 2354 | 5 Q24135 | Q24135 drosophila |
| 37 | 92.5 | 4.5 | 2354 | 5 Q9VXC8 | Q9VXC8 drosophila |
| 38 | 91.5 | 4.5 | 338 | 11 Q35400 | Q35400 mus musculu |
| 39 | 91.5 | 4.5 | 338 | 11 Q91V03 | Q91V03 mus musculu |
| 40 | 91.5 | 4.5 | 4631 | 4 Q8WXY2 | Q8WXY2 homo sapien |
| 41 | 91.5 | 4.5 | 5314 | 4 Q8WXY1 | Q8WXY1 homo sapien |
| 42 | 91.5 | 4.5 | 5938 | 4 Q96PK2 | Q96PK2 homo sapien |
| 43 | 91 | 4.5 | 294 | 11 Q9CZJ0 | Q9CZJ0 mus musculu |
| 44 | 91 | 4.5 | 1847 | 17 Q58817 | Q58817 methanococc |
| 45 | 90 | 4.4 | 573 | 5 Q9GPG0 | Q9GPG0 manduca sex |

RESULT 1

Q9Y5R3 PRELIMINARY: PRT: 386 AA.
ID Q9Y5R3;
AC Q9Y5R3;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
DE sulfotransferase GST-3).
GN GST3.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TONSIL;
RX MEDLINE=9264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin."
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TONSIL;
RX MEDLINE=9264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin."
RL J. Cell Biol. 145:899-910(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21332592; PubMed=11439191;
RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Elies L.G.,
RA Rabuka D., Hindsaugh O., Martin J.D., Lowe J.B., Fakuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT core1 extension beta1,3-N-acetylglucosaminyltransferase.";

RL Cell_105:957-969(2001).
 DR EMBL, AF133235; AAD33015.1; -;
 DR EMBL, AF280088; AAG48246.1; -;
 DR EMBL, AF149783; AAK48417.1; -;
 DR InterPro: IPR010922; HLH_basic.
 DR PROSITE: PS00038; Helix_Loop_Helix; UNKNOWN_1.
 KW lectin; Selectin; Transferase.
 SQ SQUONCE 386 AA; 45133 MW; 0C3BBA02241743A CRC64;

Query Match 100.0%; Score 2038; DB 4; Length 386;
 Best Local Similarity 100.0%; Pred. No. 1.3e-179;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPKKKLLFLVSGMATLAFEFHYSNHSLSKRAQPERHNVLYLSWRSGSSFGQ 60
 Db 1 MLLPKKKLLFLVSGMATLAFEFHYSNHSLSKRAQPERHNVLYLSWRSGSSFGQ 60
 QY 61 LFGQHPDVFYLMPEAWHVMTEFKOSTAMMLHMAVRDLRAVFLCDMSVDAVMEPPRRQ 120
 Db 61 LFGQHPDVFYLMPEAWHVMTEFKOSTAMMLHMAVRDLRAVFLCDMSVDAVMEPPRRQ 120
 QY 121 SSILFQWNSRALCSAPACDIIPODEIIPRAHCLCSQGFVEVEKACSHSVYLKEV 180
 Db 121 SSILFQWNSRALCSAPACDIIPODEIIPRAHCLCSQGFVEVEKACSHSVYLKEV 180
 QY 181 FPNLSQSLYPLLPKPSLNHLVHLVDRPRAVRSRERTKGLMIDSRIVMGQHEQKLKED 240
 Db 181 FPNLSQSLYPLLPKPSLNHLVHLVDRPRAVRSRERTKGLMIDSRIVMGQHEQKLKED 240
 QY 241 QPYVAVGVICQSOLEIYKTQSLPKAQERYLLVRYEDLARAVATSRMYEFGLEFL 300
 Db 241 QPYVAVGVICQSOLEIYKTQSLPKAQERYLLVRYEDLARAVATSRMYEFGLEFL 300
 QY 301 HLQTVWNIIRGKMGDHAFTNARALNVSQAMPWSLPEYKYSRLQKACGDAMNLGYR 360
 Db 301 HLQTVWNIIRGKMGDHAFTNARALNVSQAMPWSLPEYKYSRLQKACGDAMNLGYR 360
 QY 361 HVRSQEQRNLLDLSTWTPROI 386
 Db 361 HVRSQEQRNLLDLSTWTPROI 386

RESULT 2

Q9R111 PRELIMINARY: PRT; 388 AA.
 ID Q9R111
 AC Q9R111
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE L-selectin ligand sulfotransferase.
 GN CHST4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99361934; PubMed=10435581;
 RA Hirooka N., Petryniak B., Nakayama J., Tsudoi S., Suzuki M., Yeh J.C.,
 RA Iwawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.,
 RT "A novel, high endothelial venule-specific sulfotransferase expresses
 RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34,"
 RT Immunity 11:79-89(1999).
 DR EMBL: AF109155; AAD45579.1; -;
 DR MGD: MGI1349479; Chst4.
 KW Lectin; Selectin; Transferase.
 SQ SQUONCE 388 AA; 44636 MW; 6D5371AFB6884AEE CRC64;

Query Match 74.4%; Score 1515.5; DB 11; Length 388;
 Best Local Similarity 72.8%; Pred. No. 2e-131;
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;
 QY 1 MLLPKKKLLFLVSGMATLAFEFHYSNHSLSKRAQPERHNVLYLSWRSGSSFGV 59

Db 1 MLLPKKKLLFLVSGMATLAFEFHYSNHSLSKRAQPERHNVLYLSWRSGSSFGV 58
 QY 60 LFGQHPDVFYLMPEAWHVMTEFKOSTAMMLHMAVRDLRAVFLCDMSVDAVMEPPRR 119
 Db 59 LFGQHPDVFYLMPEAWHVMTEFKOSTAMMLHMAVRDLRAVFLCDMSVDAVMEPPRR 118
 QY 120 SSILFQWNSRALCSAPACDIIPODEIIPRAHCLCSQGFVEVEKACSHSVYLKEV 179
 Db 119 SSILFQWNSRALCSAPACDIIPODEIIPRAHCLCSQGFVEVEKACSHSVYLKEV 178
 QY 180 FPNLSQSLYPLLPKPSLNHLVHLVDRPRAVRSRERTKGLMIDSRIVMGQHEQKLKED 239
 Db 179 FPNLSQSLYPLLPKPSLNHLVHLVDRPRAVRSRERTKGLMIDSRIVMGQHEQKLKED 238
 QY 240 DQPYVAVGVICQSOLEIYKTQSLPKAQERYLLVRYEDLARAVATSRMYEFGLEFL 299
 Db 239 DQPYVAVGVICQSOLEIYKTQSLPKAQERYLLVRYEDLARAVATSRMYEFGLEFL 298
 QY 300 HLQTVWNIIRGKMGDHAFTNARALNVSQAMPWSLPEYKYSRLQKACGDAMNLGYR 359
 Db 299 HLQTVWNIIRGKMGDHAFTNARALNVSQAMPWSLPEYKYSRLQKACGDAMNLGYR 358
 QY 360 HVRSQEQRNLLDLSTWTPROI 385
 Db 359 HVRSQEQRNLLDLSTWTPROI 384

RESULT 3

Q9WUE5 PRELIMINARY: PRT; 388 AA.
 ID Q9WUE5
 AC Q9WUE5
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (Chondroitin
 DE 6/veratan) sulfotransferase 4).
 GN CHST4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=TONGUE;
 RX MEDLINE=99264336; PubMed=10330415;
 RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zhu F.-R.,
 RA Huang C.-C., Kanagaki R., Rosen S.D., Hemmerich S.,
 RT "Sulfotransferases of two specificities function in the reconstitution
 RT of high endothelial cell ligands for L-selectin,"
 RT J Cell Biol. 145:899-910(1999).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Yamana K.,
 RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batslov S., Casavant T.,
 RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Bairdrelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustinich S., Hall D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazarrelli J., Mommaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,

| | | | |
|----|-----|---|-----|
| QY | 59 | GOLGQGHEDVYLNEPAMHVMTRKOSTAMMLIHAAYDILLRAVELCMSVEFDAMIEGPR | 118 |
| Db | 57 | GOLGQGHEDVYLNEPAMHVMWDLTSGQSAAPALHAAYDILRSVFLCDMDVEFAYLT-FMR | 115 |
| QY | 119 | ROSSLFQWENRALSACPADILITQDELIPRAHCRLLCSQOPFEVYKACRSYSHVYLKE | 178 |
| Db | 116 | NISDILFQWAVRALSACPPVCEAFARGNISSEYCKPLCARPPGLAEACSSYSHVYLKE | 175 |
| QY | 179 | VRFETLQSLYLKLDPSLNLHIVLAVNDPRAVFRSRRERTGDLMSDSRIYMGHEOKLTK | 238 |
| Db | 176 | VRFETLQSLYLKLDPSLNLHIVLAVNDPRAVFRSRRERTGDLMSDSRIYMGHEOKLTK | 234 |
| QY | 239 | EDQPPYQVQVICOSOLELYK-TIOSLRKALOERILVRYEDILAPAPAOQSBRMEFEGLE | 297 |
| Db | 235 | ADPRLRVNVEVCRSHVRIAEALMKPPFTLODRIRLVREDELDAPDLTVIRELTAFFGLG | 294 |
| QY | 238 | FLPLHLQTVHNITRGKGGK--DHAFTNARDALNVSOAMRWSLPEYKVSRLQACGADM | 355 |
| Db | 235 | LTPLOLOTWITNHTGSGPGARERAEFTTTSRDALSVSOAMRWHTLPEFAKIRRVOLCGALO | 354 |
| QY | 356 | LLGGRVNRSEGOBNLLDILL 376 | |
| Db | 355 | LLGGRVNRSELEQDRLSDILL 375 | |

| | |
|--------|--|
| RESULT | 6 |
| C9GZS9 | PRELIMINARY; PRF: 411 AA. |
| ID | C9GZS9 |
| AC | C9GZS9; OSDBY3; |
| DT | 01-MAR-2001 (TrEMBLrel. 16, Created) |
| DT | 01-JUN-2001 (TrEMBLrel. 16, Last sequence update) |
| DT | 01-JUN-2002 (TrEMBLrel. 21, Last annotation update) |
| DE | Intestinal GLCNAC-6-sulfotransferase (intestinal N-acetylglucosamine-6-O-sulfotransferase) (N-acetylglucosamine 6-O-sulfotransferase). |
| GN | CHST5 OR I-GLCNAC-6-ST. |
| OS | Homo sapiens (human). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. |
| OX | NCBI_TaxId=9606; |
| RX | [1] |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=20472330; PubMed=1017086; |
| RA | Akema T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K., |
| RA | Nakamura T., Dola A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S., |
| RA | Fujiwara T., Phomar E.J., Shimomura Y., Kinoshita S., Tanigami A., |
| RA | Fukuda M.N.; |
| RT | "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene."; |
| RT | Nat. Genet. 26:237-241(2000). |
| RN | [2] |
| RP | SEQUENCE OF 22-411 FROM N.A. |
| RC | TISSUE=INTESTINE; |
| RX | MEDLINE=99423499; PubMed=10491328; |
| RA | Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.; |
| RA | "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6- |
| RL | Sulfotransferase that is Highly Restricted to Intestinal Tissue"; |
| DR | Biochem. Biophys. Res. Commun. 263:543-549(1999). |
| DR | EMBL; AF246718; FANG28023.1; - |
| DR | EMBL; AF219991; FANG26326.1; - |
| DR | EMBL; AF176839; AAD56030.1; - |
| DR | EMBL; AF176838; AAD56000.1; - |
| KW | Transference. |
| QO | SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64; |
| QY | Query Match 49.5%; Score 1008; DB 4; Length 411; Best Local Similarity 52.3%; Pred. No. 1,3e-84; Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps |
| D5 | 1 MLPEKKKKLLLEFYSSMALLALFFMHYSNNISLSMKRQPERMHVYLSSMRSGSSFVG 660 ::: ::: 35 LLAQTTCLLFITSP-----GPSSPAGCDRVAVILSSMRSGSSFTQ 800 |

| | | | |
|----|-----|---|-----|
| QY | 61 | LEGGQHPDVEYLMERPMWMMFFKOSTAMMLNARDDIIRAVFLCDMSVFPAYMEGPRQ | 139 |
| Dd | 81 | LESGHPDVEYLMERPMWMMFFKOSTAMMLNARDDIIRAVFLCDMSVFPAYMEGPRQ | 139 |
| QY | 121 | SSLFQWMSRSLCSPACADIIIPQOEIIIPRACHLLCSQPEVEYKACRSYHYLYKEVR | 180 |
| Dd | 140 | SAFEWMSRSLCSPACAPPRGTSIKDYCKLCTIROPSTLRKCRSTSYHYLYKEVR | 199 |
| QY | 181 | FFNLQSLYPLLDPSLNLCHIVLYRDPRAVRSRRETKGDIIMDSRTVMGQHEOKLEED | 240 |
| Dd | 200 | FFNLQSLYPLLDPSLNLCHIVLYRDPRAVRSRREAGPIIARONGVILGTN-GKWNEAD | 258 |
| QY | 241 | QPYVVMQYICQSOJEIYX--TQSLAPKALOEYLLVREDLAPARVQOTSMMVEVCELEF | 298 |
| Dd | 259 | PHLRITREYCSHYVIAEAALTKP-PPRLKRRYLYLVFEDLARREPLADIRALIAFTBTL | 317 |
| QY | 299 | LPHLQTVHNHITRGKMGWD--HAFHTNARDALNVSQAMRMSLPYEKYSRLQKACGDAML | 356 |
| Dd | 318 | TPOLEAMVHNHITGSGIKPTEAFHTSSRNARNVSQAMRHALPPTKILRVOEVCAGALOL | 377 |
| QY | 357 | LGRIHRESEORNLDDLL-----SIWVPE 383 | |
| Dd | 378 | LGRIHREVSANDQORDLTDLVLPRGPRDHMSMSPD 411 | |

| RESULT 7 | PRELIMINARY; | PRT; | 484 AA. |
|----------|--|------|---------|
| 09EP78 | | | |
| ID | 09EP78 | | |
| AC | 09EP78 | | |
| DT | 01-MAR-2001 (TREMBLrel. 16, Created) | | |
| DT | 01-MAR-2001 (TREMBLrel. 16, Last sequence update) | | |
| DT | 01-MAR-2002 (TREMBLrel. 20, Last annotation update) | | |
| DE | N-acetylglucosamine 6-O-sulfotransferase GST-5 (N-acetylglucosamine-6-O-sulfotransferase) (56000133070.rik protein) | | |
| DE | (Carboxylate (N-acetylglucosaminio) sulfotransferase 7). | | |
| GN | CHST7 OR GST5 OR 26000133070.rik. | | |
| OS | Mus musculus (Mouse). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| OX | NCBI_TaxID=10090. | | |
| RP | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN-C57BL/6; | | |
| RX | PubMed=10956661; | | |
| RA | Bhakta S., Bartes A., Bowman K.G., Kao W.M., Polsky I., Lee J.-K., | | |
| RA | Cook B.N., Buehl R.E., Rosen S.D., Bertozzi C.R., Hemmerlich S., | | |
| RT | "Sulfation of N-Acetylglucosamine by Chondroitin 6-Sulfotransferase 2 (GST-5)." | | |
| RL | J. Biol. Chem. 275:40226-40234 (2000). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | Uchimura K., Muramatsu T.; | | |
| RT | "Identification and molecular characterization of a cDNA encoding a novel N-acetylglucosamine-6-O-sulfotransferase." | | |
| RL | Submitted (MAR 2000) to the EMBL/GenBank/DBJ databases. | | |
| RN | [3] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN-C57BL/6J; TISSUE=EMBRYO; | | |
| RX | MEDLINE=21085660; PubMed=11217851; | | |
| RA | Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., | | |
| RA | Arkawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., | | |
| RA | Aizawa K., Iwawa M., Nishi K., Kiyosawa H., Konno S., Yamamaki I., | | |
| RA | Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R., | | |
| RA | Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., | | |
| RA | Fleischmann W., Gaasterland T., Glassi C., Kling B., Kochwa H., | | |
| RA | Kuehl P., Lewis S., Matsuo Y., Nkado I., Pesele G., Quackenbush J., | | |
| RA | Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., | | |
| RA | Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., | | |
| RA | Blake J., Boffelli D., Bojuno M., Aono H., Baldarelli R., Barsh G., | | |
| RA | Brownstein M.J., Buit C., Fletcher C., Fujita M., Gariboldi M., | | |
| RA | Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., | | |
| RA | Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombretti P., | | |
| RA | Nodone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., | | |

[illegible]

KP SEQUENCE FROM N.A.
 RC TISSUE=WHOLE EMBRYOS;
 RX MEDLINE=98380482; PubMed=9712885;
 RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
 Matsushita C., Kannagi R., Habuchi O., Muramatsu T.;
 RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
 KT sulfotransferase."
 RL J. Biol. Chem. 273:22577-22583(1998).
 DR EMBL; AB011452; BAA32138.1; -;
 DR EMBL; AB011452; BAA32139.1; -;
 DR EMBL; AB011451; BAA32137.1; -;
 DR MGD; MGI:1891160; Chs2.
 DR InterPro:IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer. 1.
 KW Transferrase
 QO SEQUENCE 350 AA; 57814 MW; A11E1B735C363EC CRC64;

[illegible]

| RESULT 13 | ID | Q9UED5 | PRELIMINARY: | PRF: | 484 AA. |
|-----------------------|--|---|--------------|-------------|---------|
| Q9UED5 | Q9UED5 | Q9UED5 | | | |
| AC | Q9UED5 | Q9UED5 | | | |
| DT | 01-MAY-2000 | (TREMBLrel. 13, Created) | | | |
| DT | 01-MAY-2000 | (TREMBLrel. 13, Last sequence update) | | | |
| DT | 01-DEC-2001 | (TREMBLrel. 19, last annotation update) | | | |
| DE | N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST). | | | | |
| GN | GNEST. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | | |
| OX | NCBI_TaxID=9606; | | | | |
| PN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | TISSUE=BRAIN; | | | | |
| RX | MEDLINE=98391845; PubMed=9722682; | | | | |
| RA | Uchiyama K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q., | | | | |
| RA | Mitsunaka C., Kanagaki R., Hahubuchi O., Yokoyama I., Yamamura K., | | | | |
| RA | Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T., | | | | |
| RT | "Human N-acetylglucosamine-6-O-sulfotransferase involved in the | | | | |
| RT | biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal | | | | |
| RL | mapping, and expression in various organs and tumor cells."; | | | | |
| DR | J. Biochem. 124:670-678(1998). | | | | |
| DR | EMBL; AB014679; BAA34265.1; | | | | |
| DR | InterPro; IPR000863; Sulfotransferase. | | | | |
| DR | Pfam; PF00685; Sulfotransfer; 1. | | | | |
| DR | Transferase. | | | | |
| SQ | SEQUENCE 484 AA; 53641 MW; D077EBCD7645F2F0 CRC64; | | | | |
| Query Match | 28.8%; | Score 587.5; | DB 4; | Length 484; | |
| Best Local Similarity | 36.0%; | Pred. No. 9.8e-46; | | | |
| Matches 131; | Conservative 67; | Mismatches 137; | Indels 29; | Gaps 6 | |
| QY | 41 ERMHVLVLSNWSGSSGSGFGOLFGCHPDDVFLYLEKPMHWMTFKOSTAWMLHMAVRLIRA | 100 | | | |
| DB | 117 KRMHWVLTFTWSSGSGFGELFNQNPVEFLYLPVHWYQKRLPGDAVSLQAGARDMLSA | 176 | | | |
| QY | 101 VFLCMSVFDAYMEBQPRRS---SLFQWNSRALCSAPACDIIFPDEIIPRAHCRLL- | 155 | | | |

| Query Match | Similarity | Score | DB | Length |
|-------------|------------|--------------|--------|----------------|
| Best Local | 36.0% | 587.5 | 29 | 484 |
| Matches | 131 | Conservative | 67 | Mismatches 137 |
| | | | Indels | 29 |
| | | | Gaps | 6 |

```

DE (GlcNAc6ST).
GN G6ST.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kanane T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsunaga C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadomatsu T., Muramatsu T.;
RT "Human N-acetylglucosamine-6-O sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal
RL J. Biochem. 124:670-678(1998).
DR EMBL: AB014680; BAA34266.1;
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransferase.
KW Sulfotransferase.
SQ
Query Match 531 AA; 56768 MW; D2EDB74E95B5162F CRC64;
Best Local Similarity 36.0%; Score 587.5; DB 4; Length 531;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMNVLVLSNRSGSSVFGQLFGQHPDVFYLMEPAMHVMWTFKOSTAMLMHAAVRLIRA 100
DB 164 KRHHMYFTTWSSGSSFFGELFNQNEVFLPEPVHVMWTKLYPGDAVSLQGAARDMLSA 223
QY 101 VFLCDMSVFAVMEPGRRROS---SLFQWNSRALCSAPACDIIPODEIIPAHORLL- 155
DB 224 LYKCDLSEVQQLYSPAGSGGNLTLLGIFGATKKNVCSPLCPAY-RKEVYGLVDRVCK 282
QY 156 -CSQDPEVEVEKACRSYSHVYLKEVRFNLQSLYPLKDPSLMHLVHLVDRPRAVFRSP 214
DB 283 KCPQRLAREFEEDCRKRTLVIGKGVFQVAVLAPLLRDALDLKYLHIVDRPRAVASSR 342
QY 215 ERTGDMIDSRIVM-----GQHEQKLRKE---DQPYV---VMQVICS 252
DB 343 IRRSHGLIRRESLOVVRSDRRAHMRPFLNAGKLGAKKKGAGVGADPADYHLAGMEVICS 402
QY 253 QLEIYKTIQSLPRALQERILLVRYEDLARAPVAGTSRMVEFGLEFLPHLOTVWHNITRG 312
DB 403 MAKLTQTLQPPDLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSPKEGECALNMTSG 462
QY 313 KGMGDHAFHFNARDALNVSGAQRMSLPYEKYSRLQKAGDANMLGVRHVRSEQQRNLL 372
DB 463 SSSSKPFVVSARNAQOANAMRTALTFOQIKQVEFCYQPMALGIERVNSPEEVKDLSS 522
QY 373 LDLL 376
DB 523 KTL 526

RESULT 15
Q9E0C0 PRELIMINARY; PRT; 411 AA.
AC Q9E0C0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Galactose 6-O-sulfotransferase GST-1.
GN GST1 OR GST1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10050;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=MAMMARY GLAND;
RA Hemmerlich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,

```

```

RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-
RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RT family.";
RT Glycobiology 11:75-87(2000).
RL EMBL: AF280087; AAG48245.1;
DR MGD; MGI:1924219; Chst1.
DR Transferrase.
KW Transferrase.
SQ
Query Match 411 AA; 46903 MW; B1AE590EF5B9CBDC CRC64;
Best Local Similarity 27.4%; Score 559; DB 11; Length 411;
Matches 132; Conservative 65; Mismatches 135; Indels 36; Gaps 10;

QY 27 YSHNLSLMAKQPERMNVLSNRSGSSVFGQLFGQHPDVFYLMEPAMHVMWTF----F 82
DB 53 FSYNLS-----RRTVLIATTRSSGSSFFGQLFNQHMVFYLFEPYHVNQTLIPRF 104
QY 83 KOSTA-----WMLHMAVRDLIRAVFLCDMSVFPVMEPGRRQSS--LFWNSRALCSAP 136
DB 105 TQKSEADRRVVLGASRDILNSLYDCDLFLENTIKPPVNHITNRVRRGASRYLCSR 164
QY 137 ACDIIP---QDEIIPRAHCRLLCSQDPEVEVEKACRSYSHVYLKEVRFNLQSLYPLKD 193
DB 165 VCD--PPGSSDLILEGDCVRMGCLLNTLAAEACRERSHVAIKTVRVENDLRALVED 222
QY 194 PSLNHLVHLVDRPRAVFRSRETKGDLMDISRIYMGQHEQKLRKEDQPYV---KQVI 249
DB 223 PLNLKVIQLVDRPGILASRSETFRDYRLMRVMYGTGR-----KPNLDVYQTLTV 275
QY 250 COSQLEIYKTIQSLPRALQERILLVRYEDLARAPVAGTSRMVEFGLEFLPHLOTVWHNI 309
DB 276 CEDPSSSVSTGLMRPSWLKGMVLRVYEDLARNPYKKEELIYFELGILDHVAHWIINN 335
QY 310 TSG-KGMGDHAFHFNARDALNVSGAQRMSLPYEKYSRLQKAGDANMLGVRHVRSEQ 368
DB 336 TEGDPTLGHKXST-VNSAATAEKWRPRLSYDIAFAQNAQOYLQAGYKMANSEEL 394
QY 369 RNLLDLL 376
DB 395 KNPALSLV 402

```

Search completed: January 11, 2003, 01:20:58
Job time : 93 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:19:36 ; Search time 24 Seconds

(without alignments)
473.219 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038
Sequence: 1 MLPRKMKLLFLVSQMAIL.....EQRLDLLDLSTWVPEQIH 386

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Parents AA:*
1: /cgn2_6/prodata/1/1aa/5a_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5b_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5a_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5b_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCRTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/beckfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|------------------|
| 1 | 2038 | 100.0 | 386 | 4 | US-09-045-284A-2 |
| 2 | 2038 | 100.0 | 386 | 4 | US-09-190-911-1 |
| 3 | 598.5 | 29.4 | 483 | 3 | US-09-263-023-2 |
| 4 | 598.5 | 29.4 | 483 | 3 | US-09-471-867-2 |
| 5 | 587.5 | 28.8 | 484 | 3 | US-09-263-023-4 |
| 6 | 587.5 | 28.8 | 484 | 4 | US-09-471-867-4 |
| 7 | 549 | 26.9 | 411 | 4 | US-09-013-188-2 |
| 8 | 500.5 | 24.6 | 458 | 2 | US-08-655-878-2 |
| 9 | 482 | 23.7 | 479 | 2 | US-08-999-514-2 |
| 10 | 94.5 | 4.6 | 848 | 4 | US-09-540-824-27 |
| 11 | 92 | 4.5 | 359 | 3 | US-09-150-133-11 |
| 12 | 92 | 4.5 | 359 | 3 | US-09-150-141-11 |
| 13 | 92 | 4.5 | 359 | 4 | US-09-374-933-11 |
| 14 | 92 | 4.5 | 359 | 4 | US-09-374-824-11 |
| 15 | 92 | 4.5 | 359 | 4 | US-09-374-933-11 |
| 16 | 87.5 | 4.3 | 380 | 3 | US-09-150-133-9 |
| 17 | 87.5 | 4.3 | 380 | 3 | US-09-150-141-9 |
| 18 | 87.5 | 4.3 | 380 | 4 | US-09-374-933-9 |
| 19 | 87.5 | 4.3 | 380 | 4 | US-09-374-824-9 |
| 20 | 87.5 | 4.3 | 380 | 4 | US-09-374-933-9 |
| 21 | 86.5 | 4.2 | 831 | 1 | US-08-073-384C-5 |
| 22 | 86.5 | 4.2 | 831 | 1 | US-08-254-355A-5 |
| 23 | 86.5 | 4.2 | 831 | 1 | US-08-483-043-5 |
| 24 | 86.5 | 4.2 | 831 | 1 | US-08-481-238-5 |
| 25 | 86.5 | 4.2 | 831 | 2 | US-08-471-066B-5 |
| 26 | 86.5 | 4.2 | 831 | 2 | US-08-484-956-5 |
| 27 | 86.5 | 4.2 | 831 | 2 | US-08-757-653-5 |

| | | | | | | |
|----|------|-----|------|---|-------------------|--------------------|
| 28 | 86.5 | 4.2 | 831 | 2 | US-08-599-491-5 | Sequence 5, Appl1 |
| 29 | 86.5 | 4.2 | 831 | 2 | US-08-756-386-5 | Sequence 5, Appl1 |
| 30 | 86.5 | 4.2 | 831 | 2 | US-08-823-516-5 | Sequence 5, Appl1 |
| 31 | 86.5 | 4.2 | 831 | 3 | US-08-682-853A-5 | Sequence 5, Appl1 |
| 32 | 86.5 | 4.2 | 831 | 3 | US-08-759-038-5 | Sequence 5, Appl1 |
| 33 | 86.5 | 4.2 | 831 | 3 | US-08-758-314-5 | Sequence 5, Appl1 |
| 34 | 86.5 | 4.2 | 831 | 3 | US-09-350-309-5 | Sequence 5, Appl1 |
| 35 | 86.5 | 4.2 | 831 | 4 | US-08-520-946-5 | Sequence 5, Appl1 |
| 36 | 83 | 4.1 | 222 | 4 | US-08-485-721-9 | Sequence 9, Appl1 |
| 37 | 83 | 4.1 | 222 | 2 | US-08-392-935-9 | Sequence 9, Appl1 |
| 38 | 83 | 4.1 | 222 | 2 | PCT-US93-08326-9 | Sequence 9, Appl1 |
| 39 | 83 | 4.1 | 920 | 3 | US-08-930-996A-8 | Sequence 8, Appl1 |
| 40 | 82.5 | 4.0 | 1140 | 4 | US-08-471-112A-4 | Sequence 4, Appl1 |
| 41 | 82.5 | 4.0 | 1809 | 3 | US-09-012-515A-12 | Sequence 12, Appl1 |
| 42 | 82.5 | 4.0 | 1809 | 4 | US-08-360-144A-12 | Sequence 12, Appl1 |
| 43 | 82.5 | 4.0 | 1809 | 4 | US-09-012-504A-12 | Sequence 3, Appl1 |
| 44 | 82.5 | 4.0 | 2549 | 4 | US-08-471-112A-3 | Sequence 12, Appl1 |
| 45 | 82.5 | 4.0 | 2549 | 5 | PCT-US95-06722-12 | Sequence 12, Appl1 |

ALIGNMENTS

RESULT 1
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/045,284A
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 100.0%; Score 2038; DB 4; Length 386;
Best Local Similarity 100.0%; Pred. No. 2, 2e-212;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | MLPRKMKLLFLVSQMAILALFFHMYSHNITSLSKAKPERKHVYLSSMRSGSFFVGO | 60 |
| DB | 1 | MLPRKMKLLFLVSQMAILALFFHMYSHNITSLSKAKPERKHVYLSSMRSGSFFVGO | 60 |
| QY | 61 | LFQGHDPVFLMEPAHWVMTFKOSTAMMLHNAVRLINAVFLCDSVFDAYMEGPFRQ | 120 |
| DB | 61 | LFQGHDPVFLMEPAHWVMTFKOSTAMMLHNAVRLINAVFLCDSVFDAYMEGPFRQ | 120 |
| QY | 121 | SSLFQWENSALCSAPACDIIPDDETIPRAHCRILCSQOPFEVVERACSYSHVYLKEVR | 180 |
| DB | 121 | SSLFQWENSALCSAPACDIIPDDETIPRAHCRILCSQOPFEVVERACSYSHVYLKEVR | 180 |
| QY | 181 | FFNLQSLYPLKDPKPSLNHLIVLRDPRAVFRSRETKGDLMDISRIYVGHQKRLKED | 240 |
| DB | 181 | FFNLQSLYPLKDPKPSLNHLIVLRDPRAVFRSRETKGDLMDISRIYVGHQKRLKED | 240 |
| QY | 241 | QPYVWQVICOQSOLEYTKITIOSLPKALQERYILVRREDILARAVPQOTSMYFVGLEFLP | 300 |
| DB | 241 | QPYVWQVICOQSOLEYTKITIOSLPKALQERYILVRREDILARAVPQOTSMYFVGLEFLP | 300 |
| QY | 301 | HLQITWENITRGKGMDAFHTNARDALNVSQAWKSLPYEKVSRLOKACGDMNLGGR | 360 |
| DB | 301 | HLQITWENITRGKGMDAFHTNARDALNVSQAWKSLPYEKVSRLOKACGDMNLGGR | 360 |
| QY | 361 | HVRSQEQRLDLLDLSTWVPEQIH 386 | |
| DB | 361 | HVRSQEQRLDLLDLSTWVPEQIH 386 | |

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

RESULT 2

US-09-190-911-1

; Sequence 1, Application US/09190911

; Patent No. 6365365

; GENERAL INFORMATION:

; APPLICANT: Bistrup, Annette

; APPLICANT: Rosen, Steven D.

; APPLICANT: Tangemann, Kirsten

; APPLICANT: Hemmerich, Stefan

; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

; FILE REFERENCE: 6510-107CIP

; CURRENT APPLICATION NUMBER: US/09/190,911

; CURRENT FILING DATE: 1998-11-12

; EARLIER APPLICATION NUMBER: 09/045,284

; EARLIER FILING DATE: 1998-03-20

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 386

; TYPE: PR1

; ORGANISM: H. sapiens

US-09-190-911-1

Query Match

Best Local Similarity 100.0%; Score 2038; DB 4; Length 386;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MLPRKKKLLFLVSCMAITLAFNNHNSHISLSKKAQPERNHVIVLSMRGSSFFVQ 60

Db 1 MLPRKKKLLFLVSCMAITLAFNNHNSHISLSKKAQPERNHVIVLSMRGSSFFVQ 60

Db 1 LFGHPVFLMEPAMVMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYNEPBRQ 120

Db 1 LFGHPVFLMEPAMVMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYNEPBRQ 120

Db 61 LFGHPVFLMEPAMVMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYNEPBRQ 120

Db 61 LFGHPVFLMEPAMVMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYNEPBRQ 120

Db 121 SFLFOWNSRALCSAPACDIIPODEIPRAHCRLLCSQCFEVEKACRSYSHVLEKVR 180

Db 121 SFLFOWNSRALCSAPACDIIPODEIPRAHCRLLCSQCFEVEKACRSYSHVLEKVR 180

Db 181 FFLNQSILYPLKPSLNLHIVLPRPRAVFRSRETKGMLTDSRIVMGOHOKLKED 240

Db 181 FFLNQSILYPLKPSLNLHIVLPRPRAVFRSRETKGMLTDSRIVMGOHOKLKED 240

Db 181 FFLNQSILYPLKPSLNLHIVLPRPRAVFRSRETKGMLTDSRIVMGOHOKLKED 240

Db 241 QPYVVMQVICOQLEIKTIOSLPKALOERYLLVRYEDLARAVAOQTSRMVEFVGLFLP 300

Db 241 QPYVVMQVICOQLEIKTIOSLPKALOERYLLVRYEDLARAVAOQTSRMVEFVGLFLP 300

Db 301 HLOTWVNITRGKMGDHAFTNARDALNVSQAMRWSLPEKYSRLQKACGDAMNLLGR 360

Db 301 HLOTWVNITRGKMGDHAFTNARDALNVSQAMRWSLPEKYSRLQKACGDAMNLLGR 360

Db 301 HLOTWVNITRGKMGDHAFTNARDALNVSQAMRWSLPEKYSRLQKACGDAMNLLGR 360

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

Db 361 HVHSEQQRNLDDLSTWTPRQIH 386

TYPE: PRT
ORGANISM: Mus musculus
US-09-471-867-2

Query Match 29.4%; Score 598.5; DB 4; Length 483;
Best Local Similarity 36.8%; Pred. No. 3e-56;
Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERMHVLSSWRSGSSFFVQGLFGQHPDYFLMEPAHVMWTFKOSTAMLMHMAVRDLIRA 100
DB KRQIVYFTTWRRSGSSFFGELFNQNPGEVFLYEPVHWVWOKLYPGDAVSLQGAARDMLSA 175
QY 101 VFLCDMSVFDAYMEPPRRROS-----SLFOWENSRALCSAPACDIIPODEIIPRAHCRLL- 155
DB LYRCDLSVFQLYSPAGSGGRNLTLLGIFGAATNKVYCCSPLCRAY-RKEVVGIVDDRVCK 234
QY 156 -CSQQPEVEVEKACRSYSHVYLKEVRFNLSQSLYPLKDPSSLNHLVHLYVRPRAVFRSR 214
DB KCPQRLAREECCRYRTLVIGVAVFVAVLAPLLRDPALDKVHLVDRBRAVASSR 294
QY 215 ERIKGLMIDSRIVM-----GOHEQKLKEDQ---DQPY---VMOVICQS 252
DB IRSRHGLIREGLQVRSRDPRAHMFLEAGHKLGAKKMGSPADYHLAGMEVICS 354
QY 253 QLEIYKTIQSLPQALQERYLLVREDLAPAPVAQTSRMVEFVGLLEFLPILOTWVHNITRG 312
DB MAKTLQTLQPPDWLQGHVLYVREDLVGDPVKTLRRYVDFVGLVSPMEQFALNMTSG 414
QY 313 KGMGDIAFHFNARDALVNSOAWKMSLPEYKVSRLQACGDANLLGYRHVRSQEQORNTL 372
DB SGGSSKPEFVVSARNAQAANAMRTALTFOOLKQVEEFCYQPAVLYGERVNSPEEVKDL 474
QY 415 373 LDLL 376
DB 475 KTL 478

RESULT 5

US-09-263-023-4
Sequence 4, Application US/09263023
Patent No. 6037159
GENERAL INFORMATION:
APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kannagi, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA1.001AUS
CURRENT APPLICATION NUMBER: US/09/263.023
CURRENT FILING DATE: 1999-03-05
EARLIER APPLICATION NUMBER: JP 10-54007
EARLIER FILING DATE: 1998-03-05
EARLIER APPLICATION NUMBER: JP 10-177844
EARLIER FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 484
TYPE: PRT
ORGANISM: Homo sapiens
US-09-263-023-4

Query Match 28.8%; Score 587.5; DB 3; Length 484;
Best Local Similarity 36.0%; Pred. No. 4.6e-55;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;
QY 41 ERMHVLSSWRSGSSFFVQGLFGQHPDYFLMEPAHVMWTFKOSTAMLMHMAVRDLIRA 100
DB KRQIVYFTTWRRSGSSFFGELFNQNPGEVFLYEPVHWVWOKLYPGDAVSLQGAARDMLSA 176

QY 101 VFLCDMSVFDAYMEPPRRROS-----SLFOWENSRALCSAPACDIIPODEIIPRAHCRLL- 155
DB LYRCDLSVFQLYSPAGSGGRNLTLLGIFGAATNKVYCCSPLCRAY-RKEVVGIVDDRVCK 235
QY 156 -CSQQPEVEVEKACRSYSHVYLKEVRFNLSQSLYPLKDPSSLNHLVHLYVRPRAVFRSR 214
DB KCPQRLAREECCRYRTLVIGVAVFVAVLAPLLRDPALDKVHLVDRBRAVASSR 295
QY 215 ERIKGLMIDSRIVM-----GOHEQKLKEDQ---DQPY---VMOVICQS 252
DB IRSRHGLIREGLQVRSRDPRAHMFLEAGHKLGAKKMGSPADYHLAGMEVICS 355
QY 253 QLEIYKTIQSLPQALQERYLLVREDLAPAPVAQTSRMVEFVGLLEFLPILOTWVHNITRG 312
DB MAKTLQTLQPPDWLQGHVLYVREDLVGDPVKTLRRYVDFVGLVSPMEQFALNMTSG 415
QY 313 KGMGDIAFHFNARDALVNSOAWKMSLPEYKVSRLQACGDANLLGYRHVRSQEQORNTL 372
DB SGGSSKPEFVVSARNAQAANAMRTALTFOOLKQVEEFCYQPAVLYGERVNSPEEVKDL 475
QY 373 LDLL 376
DB 476 KTL 479

RESULT 6

US-09-471-867-4
Sequence 4, Application US/09471867
Patent No. 6455289
GENERAL INFORMATION:
APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kannagi, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA1.001AUS
CURRENT APPLICATION NUMBER: US/09/471.867
CURRENT FILING DATE: 1999-12-23
EARLIER APPLICATION NUMBER: US 09/263.023
EARLIER FILING DATE: 1999-03-05
EARLIER APPLICATION NUMBER: JP 10-54007
EARLIER FILING DATE: 1998-03-05
EARLIER APPLICATION NUMBER: JP 10-177844
EARLIER FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 484
TYPE: PRT
ORGANISM: Homo sapiens
US-09-471-867-4

Query Match 28.8%; Score 587.5; DB 4; Length 484;
Best Local Similarity 36.0%; Pred. No. 4.6e-55;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;
QY 41 ERMHVLSSWRSGSSFFVQGLFGQHPDYFLMEPAHVMWTFKOSTAMLMHMAVRDLIRA 100
DB KRQIVYFTTWRRSGSSFFGELFNQNPGEVFLYEPVHWVWOKLYPGDAVSLQGAARDMLSA 176
QY 101 VFLCDMSVFDAYMEPPRRROS-----SLFOWENSRALCSAPACDIIPODEIIPRAHCRLL- 155
DB LYRCDLSVFQLYSPAGSGGRNLTLLGIFGAATNKVYCCSPLCRAY-RKEVVGIVDDRVCK 235
QY 156 -CSQQPEVEVEKACRSYSHVYLKEVRFNLSQSLYPLKDPSSLNHLVHLYVRPRAVFRSR 214
DB KCPQRLAREECCRYRTLVIGVAVFVAVLAPLLRDPALDKVHLVDRBRAVASSR 295
QY 215 ERIKGLMIDSRIVM-----GOHEQKLKEDQ---DQPY---VMOVICQS 252
DB KCPQRLAREECCRYRTLVIGVAVFVAVLAPLLRDPALDKVHLVDRBRAVASSR 295

```
Db 296 IRSRHGLRESLOVRSRDRRAHMPFLBAGHKLGAKEGVGADYHALGAMETICNS 355
Qy 253 QLEIKYKISGLPALDERYLIVYEDLARPAVQTSRMEEFVLEFLPHIQTVHNITG 312
Db 356 MXTLQTLQPLPQWMLQGHVLYREDLVGDVYKTLRRVYDEVGLVSPMEQPALMWTG 415
Qy 313 KGMGDHAFHTNARDALVNSQAMWMSLPEYKVSRLQKACGDAMNLGGRHVRSQEQRNLL 372
Db 416 SGSSSKRFYVSAHNATQGAAMAMTALTFOQIKQVEEFCQPMAYLVGRVNSPEYKDL 475
Qy 373 LDLL 376
Db 476 KTL 479

RESULT 7
US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroctin
; TIME OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2

Query Match 26.9%; Score 549; DB 4; Length 411;
Best Local Similarity 34.3%; Pred No. 5.4e-51;
Matches 139; Conservative 64; Mismatches 160; Indels 42; Gaps 10;

Qy 6 KMKLLFLVS--OMALLAFHMYSHNITSLSMKAQPERM-----HVLV 47
Db 6 KAVLLLASIAIQYTAIRFTAKSFHTCPGLAELAEELCESEPTFAYNLSRKTIL 65
Qy 48 LSSWRSSSVFQGLFGHDPVFLMEPAHVMWT-----FKQSTA-----MMLNVAVDLLR 99
Db 66 LATTRSSSVFQGLFQNDLVFLFEPLVHVQNTLLPFTQKSPADRRVMGASRDLLR 125
Qy 100 AFLVCDMSVDAVMEGPRROSS--LEQWENSRLCSAPACDII--PQDEIIPRAHCLIC 156
Db 126 SLVDCDLYFLENYIKPPVNHHTDRIEFGASVILCSRPVCDPPADLVLEGGDCVRC 185
Qy 157 SQCPFEYVERACRSYSHVYLKEYRFENLOSPLKDPSLNHTIYLVDRPRAVFNRSER 216
Db 186 GLNLTVAAACRERSHVALKTVRPEVNDLRALVEDPRLNLKVIQLVDRPGLIASRSE 245
Qy 217 TKGLMDSIRIVGQHQKLEKEDOPYV---MOVICQSLEIYKISGLPALDERYL 272
Db 246 TFRRTYLMWLMWGTG-----KRYNDVYQLTVCEDEFSNSYSTGLMRPWLKGYKN 298
Qy 273 LVREYEDARPAVQTSRMEEFVLEFLPHIQTVHNITG--KMGDHAFTNARDALVNS 331
Db 299 LVREYEDARPAVQTSRMEEFVLEFLPHIQTVHNITG--KMGDHAFTNARDALVNS 357
Qy 332 QANRWSLPEYKVSRLQKACGDAMNLGGRHVRSQEQRNLLDL 376
Db 358 EKMFRLSLYIDVAFPAQVACQVLAQLGYKIASSEELKNPSVSLV 402

RESULT 8
US-08-655-878-2
; Sequence 2, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
```

```
APPLICANT: FUKUTA, MASAKAZU
APPLICANT: HABUCHI, OSAMI
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS:
STREET:
CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 458
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-655-878-2

Query Match 24.6%; Score 500.5; DB 2; Length 458;
Best Local Similarity 33.2%; Pred. No. 1.1e-45;
Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

Qy 34 LSMKAQPERMAYLVSSWRSSVFQGLFGHDPVFLMEPAHVM--WMTEKQSTAMWLH 91
Db 104 LGIAAPERRRHVLLMATRTGSSFYEGEFNOGNIFFLEPLMIIEKRVYTFEPGANAAG 163
Qy 92 MAV--RDLIRAVFLCDMSVDAVMEGPRRO--SLEQWENSRLCSAPACDIIPO--DEI 146
Db 164 SALVYRDVQLQLLDCDLYLLESFISPADEEHLLTALFRGSSHSLSCEPVC--TPSLKY 221
Qy 147 IPRAHCR--LLSQCPFEYVERACRSYSHVYLKEYRFENLOSPLKDPSLNHTIYLVDR 205
Db 222 FEKYCKNRRCGPLNITLAEACRKRQHMALKTVIRIROLEFLQPLAEDPRLDLTIQLVR 281
Qy 206 DPRAVFRSERTGDLMDISRIVQHQKLEK-----KEDQPYVVMQVICS--QLE 255
Db 282 DPRAVLVR-----AV---AFSGKESYKWKMAEGEAPQDEE--VQRLNGCESIRLS 330
Qy 256 IYKTIQSLPKALOETYLIVREYEDLARPAVQTSRMEEFVLEFLPHIQTVHNITGKGM 315
Db 331 AELGRQ--PRMLRGHYMLVREYEDVARAPLRKALEYRPAIGHPYQVEWIRANTQAP--Q 388
Qy 316 GDHAFHTNARDALVNSQAMWMSLPEYKVSRLQKACGDAMNLGGRHVRSQEQRNLLDL 375
Db 389 DSNGLYTSQKNSSEQFEKMRRESIPFKLAQVYODACERPMRLFGYKILASSAQELTNRSL 448
Qy 376 L 376
Db 449 L 449

RESULT 9
US-08-899-514-2
```


Mon Jan 13 09:14:05 2003

us-09-816-825-2.ra1

Page 5

Sequence 2 Application US/08839514
Patent No. 5910581
GENERAL INFORMATION:
APPLICANT: HAEUCHI, OSAMI
APPLICANT: FUKUDA, MASAKAZU
TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
TITLE OF INVENTION: SUBOTRANSFERASE ORIGINATING FROM HUWAN AND DNA CODING
TITLE OF INVENTION: FOR THE POLYPEPTIDE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBE, MARTENS, OLSON & BEAR, LLP
STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
CITY: NEWPORT BEACH
STATE: CALIFORNIA
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,514
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DANIEL E ALTMAN
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: TOWAM21.001AUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 479
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-899-514-2

```

QY 108 VFDAYMEPRGRROSSLFGWENSALCSAPACDIIPQ-DEIIPRAHCRLLCSQOFVEVER 166
| : : : : : | : : : : | : : : : |
Db 144 VFDDAV-----SATTETIVAKHSELAPR-----LCNKDP----- 172
QY 167 ACSYSHVVLKEVREFNLSLYPLKDPSSLNLIHVLVDRPRAVRSRRTK---GDLN 222
| : : : : : | : : : : | : : : : |
Db 173 ---YALMLPTIR-----RLYP-----NAKFLIMIRDARAVHSMIERKVPVAGYNT 216
QY 223 IDSRIWGOHEQKLKEDQPYVMQVICOQO---LEIKTIOSLPKALOERLLVRYED 278
| : : : : : | : : : : | : : : : |
Db 217 SDEISMVQWNOELRK-----MTFOCNNAPOGQCIKYV-----YER 251
QY 279 IARAPVQTSRMVEFVGLFELPHLOTWVHNITRGKMGDAFH-TNARDALNVSOAWRWS 337
| : : : : : | : : : : | : : : : |
Db 252 LIQKPAEILIRITNFDLPFSQOMLRHODLIGDEVLDNDQEFASQVKNKSNITKALTSWF 311
QY 338 LPY-EKVSRLQKACGDAMNLLGY 359
| : : : : : | : : : : |
Db 312 DCFSEETLRKLDVAPFLGILGY 334

RESULT 12
US-09-150-141-11
; Sequence 11, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.493
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-141-11

Query Match 4.5%; Score 92; DB 3; Length 359;
Best Local Similarity 19.5%; Pred. No. 0.14;
Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;

QY 52 RSGSSEFVQQLFGQHPDV---FYLMPEAWHYWMTFKOSTAMMLHMAVRDLIRAVFLCDMS 107
| : : : : : | : : : : | : : : : |
Db 95 RSGTTLRRAILDHAPDYRCGGETMLDPSFLTWQAGWRND-WVNNSGI-----TQE 143
QY 108 VFDAYMEPRGRROSSLFGWENSALCSAPACDIIPQ-DEIIPRAHCRLLCSQOFVEVER 166
| : : : : : | : : : : | : : : : |
Db 144 VFDDAV-----SATTETIVAKHSELAPR-----LCNKDP----- 172
QY 167 ACSYSHVVLKEVREFNLSLYPLKDPSSLNLIHVLVDRPRAVRSRRTK---GDLN 222
| : : : : : | : : : : | : : : : |
Db 173 ---YALMLPTIR-----RLYP-----NAKFLIMIRDARAVHSMIERKVPVAGYNT 216
QY 223 IDSRIWGOHEQKLKEDQPYVMQVICOQO---LEIKTIOSLPKALOERLLVRYED 278
| : : : : : | : : : : | : : : : |
Db 217 SDEISMVQWNOELRK-----MTFOCNNAPOGQCIKYV-----YER 251
QY 279 IARAPVQTSRMVEFVGLFELPHLOTWVHNITRGKMGDAFH-TNARDALNVSOAWRWS 337
| : : : : : | : : : : | : : : : |
Db 252 LIQKPAEILIRITNFDLPFSQOMLRHODLIGDEVLDNDQEFASQVKNKSNITKALTSWF 311
QY 338 LPY-EKVSRLQKACGDAMNLLGY 359
| : : : : : | : : : : |
Db 312 DCFSEETLRKLDVAPFLGILGY 334

RESULT 13
US-09-374-493-11
; Sequence 11, Application US/09374493

```

```

; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-493-11

Query Match 4.5%; Score 92; DB 4; Length 359;
Best Local Similarity 19.5%; Pred. No. 0.14;
Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;

QY 52 RSGSSEFVQQLFGQHPDV---FYLMPEAWHYWMTFKOSTAMMLHMAVRDLIRAVFLCDMS 107
| : : : : : | : : : : | : : : : |
Db 95 RSGTTLRRAILDHAPDYRCGGETMLDPSFLTWQAGWRND-WVNNSGI-----TQE 143
QY 108 VFDAYMEPRGRROSSLFGWENSALCSAPACDIIPQ-DEIIPRAHCRLLCSQOFVEVER 166
| : : : : : | : : : : | : : : : |
Db 144 VFDDAV-----SATTETIVAKHSELAPR-----LCNKDP----- 172
QY 167 ACSYSHVVLKEVREFNLSLYPLKDPSSLNLIHVLVDRPRAVRSRRTK---GDLN 222
| : : : : : | : : : : | : : : : |
Db 173 ---YALMLPTIR-----RLYP-----NAKFLIMIRDARAVHSMIERKVPVAGYNT 216
QY 223 IDSRIWGOHEQKLKEDQPYVMQVICOQO---LEIKTIOSLPKALOERLLVRYED 278
| : : : : : | : : : : | : : : : |
Db 217 SDEISMVQWNOELRK-----MTFOCNNAPOGQCIKYV-----YER 251
QY 279 IARAPVQTSRMVEFVGLFELPHLOTWVHNITRGKMGDAFH-TNARDALNVSOAWRWS 337
| : : : : : | : : : : | : : : : |
Db 252 LIQKPAEILIRITNFDLPFSQOMLRHODLIGDEVLDNDQEFASQVKNKSNITKALTSWF 311
QY 338 LPY-EKVSRLQKACGDAMNLLGY 359
| : : : : : | : : : : |
Db 312 DCFSEETLRKLDVAPFLGILGY 334

RESULT 14
US-09-374-824-11
; Sequence 11, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-824-11

```

Mon Jan 13 09:14:05 2003

us-09-816-825-2.ra1

```
Query Match          4.5%; Score 92; DB 4; Length 359;
Best Local Similarity 19.5%; Pred. No. 0.14;
Matches 63; Conservative 46; Mismatches 114; Indels 98; Gaps 15;

QY 52 RSGSSFYGOLFQHPDY----FYIMEPAMHVMTFKOSTAMLMAMAYRDLIRAVFLCDMS 107
   |||: : |||: : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 RSGTTLRAILDAHPDVRCGETWLLPSFLTWQAGWRND-WVNNSGI-----TQE 143

QY 108 VFDAVMEPPRRQSLQWENSRLCSAPACDIIPQ-DEIIPRAHCLLCSQOPFEVVEK 166
   |||: : |||: : : : : : : : : : : : : : : : : : : : : : : : :
Db 144 VFDDAV-----SAFITEIVAKHSELAPR-----LCKNDP----- 172

QY 167 ACRSYSHVLKEVFENLQSLYPLKDPISLNIHIVLVRDPRAVFRSRETK----GDLN 222
   |||: : |||: : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 ----YTLMLPTIR-----RLYP-----NAKFIIMIRARAVHSMIERKVPVAGYNT 216

QY 223 IDSRIWGOEHQKLKEDQPIYWMQVICSQ---LEIYKTIOSLPKALQERYILVRED 278
   |||: : |||: : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 SDEISMFEVQNMQELRK-----MTFOCNNAPGQCICKVY-----YER 251

QY 279 LARAPVQTSRMKEFVGLEFLPHLQTWVHNITRKGKMGDAFH-TNARDALNVSQAMRMS 337
   |||: : |||: : : : : : : : : : : : : : : : : : : : : : : : :
Db 252 LIQKPAEILIRITNFDLPFSQQMLRHODLIGDEVLDNDQEFSSAQVKNISINTKALTSMF 311

QY 338 LPY-EKVSRLQKACGDAMNLGY 359
   |||: : |||: : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 DCFSEETLRKLDVAPFLGITGY 334

RESULT 15
US-09-374-492-11
; Sequence 11, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 3820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-492-11

Query Match          4.5%; Score 92; DB 4; Length 359;
Best Local Similarity 19.5%; Pred. No. 0.14;
Matches 63; Conservative 46; Mismatches 114; Indels 98; Gaps 15;
```

```
QY 279 LARAPVQTSRMKEFVGLEFLPHLQTWVHNITRKGKMGDAFH-TNARDALNVSQAMRMS 337
   |||: : |||: : : : : : : : : : : : : : : : : : : : : : : : :
Db 252 LIQKPAEILIRITNFDLPFSQQMLRHODLIGDEVLDNDQEFSSAQVKNISINTKALTSMF 311

QY 338 LPY-EKVSRLQKACGDAMNLGY 359
   |||: : |||: : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 DCFSEETLRKLDVAPFLGITGY 334
```

Search completed: January 11, 2003, 01:22:24
Job time : 28 secs

This Page Blank (uspto)

Mon Jan 13 09:14:06 2003

us-09-816-825-2.rapb

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:21:07 ; Search time 18 seconds

(without alignments)
416.045 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038

Sequence: 1 MLPPKMKKLLFLVSQMAI.....EQRLNLLDLSTWVEQIH 386

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----------------------|---------------------|
| 1 | 2038 | 100.0 | 386 | 9 US-10-007-262-1 | Sequence 1, Appl 1 |
| 2 | 2038 | 100.0 | 386 | 10 US-09-816-825-2 | Sequence 2, Appl 1 |
| 3 | 1028 | 50.4 | 418 | 10 US-09-927-602-5 | Sequence 5, Appl 1 |
| 4 | 1019 | 50.0 | 395 | 10 US-09-927-602-2 | Sequence 2, Appl 1 |
| 5 | 1013.5 | 49.7 | 395 | 10 US-09-927-602-3 | Sequence 3, Appl 1 |
| 6 | 1008 | 49.5 | 390 | 10 US-09-827-602-4 | Sequence 4, Appl 1 |
| 7 | 821 | 40.3 | 171 | 10 US-09-827-602-8 | Sequence 8, Appl 1 |
| 8 | 587.5 | 28.8 | 531 | 10 US-09-833-790-255 | Sequence 255, App |
| 9 | 519.5 | 25.5 | 169 | 10 US-09-927-602-6 | Sequence 6, Appl 1 |
| 10 | 511.5 | 25.1 | 169 | 10 US-09-927-602-7 | Sequence 7, Appl 1 |
| 11 | 336 | 16.5 | 179 | 10 US-09-927-602-9 | Sequence 9, Appl 1 |
| 12 | 335.5 | 16.5 | 179 | 10 US-09-927-602-10 | Sequence 10, Appl 1 |
| 13 | 293 | 14.4 | 174 | 10 US-09-827-602-11 | Sequence 11, Appl 1 |
| 14 | 83 | 4.4 | 320 | 10 US-09-854-122-20 | Sequence 20, Appl 1 |
| 15 | 88 | 4.3 | 2209 | 9 US-09-902-941-1903 | Sequence 1903, Ap |
| 16 | 88 | 4.3 | 2209 | 9 US-09-849-626-1903 | Sequence 1903, Ap |
| 17 | 86.5 | 4.2 | 831 | 9 US-10-033-297-5 | Sequence 5, Appl 1 |
| 18 | 86.5 | 4.2 | 831 | 9 US-10-081-806-5 | Sequence 5, Appl 1 |
| 19 | 86 | 4.2 | 824 | 10 US-09-801-368-312 | Sequence 312, App |

| | | | | | |
|----|------|-----|------|------------------------|---------------------|
| 20 | 81 | 4.0 | 295 | 9 US-09-965-529-31 | Sequence 31, Appl 1 |
| 21 | 81 | 4.0 | 1100 | 10 US-09-815-242-11861 | Sequence 11861, A |
| 22 | 79.5 | 3.9 | 177 | 8 US-08-424-5508-53 | Sequence 53, Appl 1 |
| 23 | 79 | 3.9 | 402 | 10 US-09-815-242-11350 | Sequence 11350, A |
| 24 | 78.5 | 3.9 | 476 | 9 US-09-925-300-1612 | Sequence 1612, Ap |
| 25 | 78.5 | 3.9 | 745 | 10 US-09-844-988-10 | Sequence 10, Appl 1 |
| 26 | 78.5 | 3.9 | 745 | 10 US-09-796-872-2 | Sequence 2, Appl 1 |
| 27 | 78.5 | 3.9 | 745 | 10 US-09-844-908-10 | Sequence 12522, A |
| 28 | 78.5 | 3.9 | 899 | 10 US-09-815-242-13556 | Sequence 5356, Ap |
| 29 | 78.5 | 3.9 | 992 | 10 US-09-815-242-12615 | Sequence 12615, A |
| 30 | 78 | 3.8 | 244 | 10 US-09-911-826A-9 | Sequence 9, Appl 1 |
| 31 | 78 | 3.8 | 1160 | 10 US-09-815-242-5480 | Sequence 5480, Ap |
| 32 | 77.5 | 3.8 | 1168 | 10 US-09-815-242-12151 | Sequence 12151, A |
| 33 | 77.5 | 3.8 | 1127 | 10 US-09-815-242-5373 | Sequence 5373, Ap |
| 34 | 77.5 | 3.8 | 1158 | 10 US-09-815-242-12522 | Sequence 12522, A |
| 35 | 77 | 3.8 | 575 | 10 US-09-815-242-10240 | Sequence 10240, A |
| 36 | 76 | 3.7 | 500 | 10 US-09-731-872-466 | Sequence 466, App |
| 37 | 75.5 | 3.7 | 748 | 10 US-09-801-574-40 | Sequence 40, Appl 1 |
| 38 | 75.5 | 3.7 | 1285 | 10 US-09-982-091A-2 | Sequence 2, Appl 1 |
| 39 | 75 | 3.7 | 1179 | 10 US-09-815-242-13608 | Sequence 13608, A |
| 40 | 74.5 | 3.7 | 666 | 9 US-09-331-631A-3 | Sequence 3, Appl 1 |
| 41 | 74.5 | 3.7 | 755 | 10 US-09-881-852-19 | Sequence 19, Appl 1 |
| 42 | 74 | 3.6 | 538 | 9 US-09-991-496-118 | Sequence 118, App |
| 43 | 74 | 3.6 | 538 | 10 US-09-874-923-118 | Sequence 118, App |
| 44 | 74 | 3.6 | 559 | 9 US-09-895-913A-304 | Sequence 304, App |
| 45 | 74 | 3.6 | 628 | 10 US-09-881-752A-144 | Sequence 144, App |

ALIGNMENTS

| | | | | | | | | | |
|--|-----|--|--------------|-----------|---------|----|--|--|--|
| RESULT 1 | | | | | | | | | |
| US-10-007-262-1 | | | | | | | | | |
| ; Sequence 1, Application US/10007262 | | | | | | | | | |
| ; Patent No. US20020164748A1 | | | | | | | | | |
| GENERAL INFORMATION: | | | | | | | | | |
| ; APPLICANT: Bistrup, Annette | | | | | | | | | |
| ; APPLICANT: Rosen, Steven D. | | | | | | | | | |
| ; APPLICANT: Tangemann, Kirsten | | | | | | | | | |
| ; APPLICANT: Hemmerlich, Stefan | | | | | | | | | |
| ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3 | | | | | | | | | |
| ; FILE REFERENCE: 6510-107CIP | | | | | | | | | |
| ; CURRENT APPLICATION NUMBER: US/10/007,262 | | | | | | | | | |
| ; CURRENT FILING DATE: 2001-11-08 | | | | | | | | | |
| ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911 | | | | | | | | | |
| ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12 | | | | | | | | | |
| ; NUMBER OF SEQ ID NOS: 8 | | | | | | | | | |
| ; SOFTWARE: FastSeq for Windows Version 3.0 | | | | | | | | | |
| ; SEQ ID NO 1 | | | | | | | | | |
| ; LENGTH: 386 | | | | | | | | | |
| ; TYPE: PRT | | | | | | | | | |
| ; ORGANISM: H. sapiens | | | | | | | | | |
| US-10-007-262-1 | | | | | | | | | |
| Query Match | | | | | | | | | |
| Best Local Similarity 100.0%; Score 2038; DB 9; Length 386; | | | | | | | | | |
| Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0 | | | | | | | | | |
| QY | 1 | MLPPKMKKLLFLVSOMAILALFFHMYSHN | ISLSMKACPERM | HVLVLSWRS | GSSSFVQ | 60 | | | |
| Db | 1 | MLPPKMKKLLFLVSOMAILALFFHMYSHN | ISLSMKACPERM | HVLVLSWRS | GSSSFVQ | 60 | | | |
| QY | 61 | LFQGHPPVFILMPAMVWMTFKOSTAMLMHVPDLIRAVFLCDMSVDPAYNEP | PPRQ | 120 | | | | | |
| Db | 61 | LFQGHPPVFILMPAMVWMTFKOSTAMLMHVPDLIRAVFLCDMSVDPAYNEP | PPRQ | 120 | | | | | |
| QY | 121 | SSLFQWNSRALCSAPACDIIPODEIIPRAHCRLLCSQGFVEYERAKCSYSHV | LKEVR | 180 | | | | | |
| Db | 121 | SSLFQWNSRALCSAPACDIIPODEIIPRAHCRLLCSQGFVEYERAKCSYSHV | LKEVR | 180 | | | | | |
| QY | 181 | FFNQSILYPLKDPSTLNHLIVLPDRAVFRSRETKGMDSDRITVWGQHEOKLKED | 240 | | | | | | |
| Db | 181 | FFNQSILYPLKDPSTLNHLIVLPDRAVFRSRETKGMDSDRITVWGQHEOKLKED | 240 | | | | | | |


```

RESULT 13
US-09-927-602-11
; Sequence 11, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-17 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-11

Query Match          14.4%: Score 293; DB 10; Length 174;
Best Local Similarity 25.6%: Pred. No. 7.9e-22;
Matches 69; Conservative 30; Mismatches 59; Indels 112; Gaps 2.

YQ  42 RMHYLVSSRSGSSVGVGLFGGHDPVYLVKEPAHWYMTFKOSTAMWLMHAYDLIRAY 101
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  12 RRVHLMAITRTGSSVGVGEFFNQGNITPLPEPLMHIRITS----- 53

YQ  102 FLCDMSVFADYMEPGFRGSSSLFQWENSBALCSAPACDIIPDEIIPRAHRLCSQOP 161
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  54 ----- 53

YQ  162 EYERKACRSYSHVYLKEVRFENFNLQSLPLKDPSSLNTHIVLVDPRAVRSRRTKGDL 221
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  54 --FEARCRREHMLAAVRIROLEFLQPLADPRDLVLIQVADPRAYLASR----- 104

YQ  222 MDSRIWVGQHEKRLKEDQPIYVMQVICOSELYIKTQSLPKALOERVLVRYEDLAR 281
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  105 -----WVAFAAGKYKTMELGLRQPAWMLRGVRLVRYEDVAR 139

YQ  282 APVAQTSRMAYEYGLFPLHLQTWYHNTR 311
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  140 GFLQKAREMYPFAGIPLTPQVEDWIKXNQ 169

RESULT 14
US-09-854-122-20
; Sequence 20, Application US/09854122
; Patent No. US20020016980A1
; GENERAL INFORMATION:
; APPLICANT: ALBERT, RANDALL S.
; APPLICANT: SMITH, ROBERT
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA
; FILE REFERENCE: PHA-007.01
; CURRENT APPLICATION NUMBER: US/09/854,122
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/202,529
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Flaveria bidentis
US-09-854-122-20

Query Match          4.4%: Score 89; DB 10; Length 320;
Best Local Similarity 20.8%: Pred. No. 0.39;
Matches 56; Conservative 30; Mismatches 83; Indels 100; Gaps 12

```

| QY | 54 | GSSTVGQFLGQGH-DVEFLMEPAHVMHMTFKQSTAPMLHMAVRDLIRAVFLCDMSVFDAY | 112 |
|--|------|--|------|
| Db | 47 | GALLAQGSKRARPDDVFLCSTP-----KSGITWLKALYAVTR-----EKDEDF | 91 |
| QY | 113 | MEP-----GGRQSSLFQ-WENSALCSAPACDIIPODEITPRAHCRLLCSQOPE | 162 |
| Db | 92 | TSPFLTINPHNCIPETIEKDLKKRIVENQNSCFTP-----MATHMPHY | 133 |
| QY | 163 | VVEACRSYSHVVLKEVAFNVLQSLIYPLLKPSLNIHIVLVPDPAVFRSERTKGLM | 222 |
| Db | 134 | VLPFS-----TL-----ALMCKNVIYIRNKDY----- | 156 |
| QY | 223 | IDSRIWGMGHEOKLTKKEDQPYVMQVICOQSOLEIKYITQSLP-----KALOER-- | 270 |
| Db | 157 | IVSEHFHREITRKLPLEDPAPE-----EAEDEFHGICISQGPYWDHLLGYMKASLERPE | 210 |
| QY | 271 | YLLVREEDLARAPVAQTSRMVEFVGLTF | 298 |
| Db | 211 | VILFLKYEVRKDPSTSNYKRLAEFLGTFYF | 239 |
| RESULT 15 | | | |
| US-09-902-941-1903 | | | |
| Sequence 1903, Application us/09902941 | | | |
| Patent No. US20020172952A1 | | | |
| GENERAL INFORMATION: | | | |
| APPLICANT: Henderson, Robert A. | | | |
| APPLICANT: Wang, Tonglong | | | |
| APPLICANT: Watanabe, Yoshihiro | | | |
| APPLICANT: Johnson, Jeffrey C. | | | |
| APPLICANT: Retter, Marc W. | | | |
| APPLICANT: Marnetakis, Margarita | | | |
| APPLICANT: Carter, Derrick | | | |
| APPLICANT: Ranger, Gary R. | | | |
| APPLICANT: Vedvick, Thomas S. | | | |
| APPLICANT: Bangur, Chaitanya S. | | | |
| APPLICANT: Mcnabb, Andria | | | |
| TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY | | | |
| TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER | | | |
| FILE REFERENCE: 210121.478C17 | | | |
| CURRENT FILING DATE: 2001-07-10 | | | |
| NUMBER OF SEQ ID NOS: 2002 | | | |
| SOFTWARE: FASTSEQ for Windows Version 4.0 | | | |
| SEQ ID NO 1903 | | | |
| LENGTH: 2209 | | | |
| TYPE: PRT | | | |
| ORGANISM: Homo sapiens | | | |
| US-09-902-941-1903 | | | |
| Query Match | | | |
| Best Local Similarity 4.3%, Score 88; DB 9; Length 2209; | | | |
| Matches 54; Conservative 48; Mismatches 98; Indels 134; Gaps 10; | | | |
| QY | 29 | HNISSTSMKACQERHNVILVSSMRGSSGVGLFG---QHPVFL-----M | 72 |
| Db | 1253 | NSISHLLOWLQSS-----SOMELALRRVVGSEFGICQHSVSNMNTLSKELFGETTL | 1305 |
| QY | 73 | EPAAHVMTFKOSTAMMLHMAVRDLIRAVFLC---DMSYFDAYMPEGRPRQSSLTQWENS | 129 |
| Db | 1306 | VSRHRHVMELKRAVAFIRENAFTLLHKYFNCRLVLDLALCY----- | 1348 |
| QY | 130 | RALCSAPACDIIPODEITPRAHCRLLCSQOPEVEVEKACRSYSHVVLKEVREFNLOSTY- | 188 |
| Db | 1349 | -----CTLPQNDYF-----ENMKLLDKAMQNDKILALSLVSGSELASTIQ | 1390 |
| QY | 189 | -----PLKPSLNLHIVLVDPAVFRSERTKGLM-----IDSRIWMO | 231 |
| Db | 1391 | ETEMGLKREFLESTDAQMGIRLGLGTSFQVPRQHFLTKDKILALVENIDWDSTIL-- | 1448 |
| QY | 232 | HQGLKKEPQPYVMQVICOQSOLEIKYITQSLP-KALOERYLLVREEDLARAPVAQTSRM | 291 |

Mon Jan 13 09:14:06 2003

us-09-816-825-2.rapb

Page 7

Db 1449 -----EXCSTFQJDCDVAVLQJF----- 1465

QY 292 EFVGLFPLPHJQFWVHNITRGKSGDHAFHTNAR 325

Db 1466 -----IETLHNTNAGQGQGDASMSAKR 1489

Search completed: January 11, 2003, 01:22:57
Job time : 22 secs

This Page Blank (uspto)

Mon Jan 13 09:13:55 2003

us-09-816-825-1.ol120.rge

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 21:40:22 ; Search time 5288 Seconds

(without alignments)
11183.226 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032

Sequence: 1 ggcctcagagccagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 2054640 segs, 14551402878 residues

Word size: 20

Total number of hits satisfying chosen parameters: 1438

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_cv:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pla:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vtc:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 2032 | 100.0 | 2032 | 6 AR203335 | AR203335 Sequence |
| 2 | 2032 | 100.0 | 2032 | 9 AF131235 | AF131235 Homo sapi |
| 3 | 1856 | 91.3 | 2011 | 9 AK026635 | AK026635 Homo sapi |
| 4 | 1806 | 88.9 | 183228 | 9 AC010547 | AC010547 Homo sapi |
| 5 | 1756 | 86.4 | 1992 | 9 AF280088 | AF280088 Homo sapi |
| 6 | 1211 | 59.6 | 1333 | 9 AF149783 | AF149783 Homo sapi |
| 7 | 505 | 24.9 | 517 | 6 AX381256 | AX381256 Sequence |
| 8 | 266 | 14.1 | 389 | 6 AX381326 | AX381326 Sequence |
| 9 | 259 | 2.9 | 1462 | 9 AF176839 | AF176839 Homo sapi |
| 10 | 59 | 2.9 | 1647 | 6 AX327330 | AX327330 Sequence |
| 11 | 59 | 2.9 | 2170 | 9 AF176838 | AF176838 Homo sapi |
| 12 | 59 | 2.9 | 2544 | 9 AF219990 | AF219990 Homo sapi |
| 13 | 59 | 2.9 | 3278 | 9 AF246718 | AF246718 Homo sapi |
| 14 | 59 | 2.9 | 3786 | 9 AF280086 | AF280086 Homo sapi |
| 15 | 59 | 2.9 | 71503 | 9 AF219991 | AF219991 Homo sapi |
| 16 | 59 | 2.9 | 157337 | 9 AC009163 | AC009163 Homo sapi |
| 17 | 59 | 2.9 | 194832 | 9 AC025287 | AC025287 Homo sapi |
| 18 | 59 | 2.9 | 208185 | 2 AC009105 | AC009105 Homo sapi |
| 19 | 37 | 1.8 | 37 | 6 AR203340 | AR203340 Sequence |
| 20 | 27 | 1.3 | 1926 | 10 AF131236 | AF131236 Mus muscu |
| 21 | 27 | 1.3 | 2201 | 10 AF109155 | AF109155 Mus muscu |
| 22 | 24 | 1.2 | 1398 | 9 AF302109 | AF302109 Homo sapi |
| 23 | 24 | 1.2 | 79388 | 2 AC012281 | AC012281 Homo sapi |
| 24 | 24 | 1.2 | 99395 | 2 AC095651 | AC095651 Rattus no |
| 25 | 24 | 1.2 | 131969 | 2 AC116944 | AC116944 Tetradon |
| 26 | 24 | 1.2 | 143655 | 2 AC010808 | AC010808 Homo sapi |
| 27 | 24 | 1.2 | 168991 | 2 AC073958 | AC073958 Homo sapi |
| 28 | 24 | 1.2 | 181557 | 2 AC072055 | AC072055 Homo sapi |
| 29 | 24 | 1.2 | 191639 | 2 AC116046 | AC116046 Papio ham |
| 30 | 23 | 1.1 | 1740 | 10 AF176841 | AF176841 Mus muscu |
| 31 | 23 | 1.1 | 1989 | 10 AF176840 | AF176840 Mus muscu |
| 32 | 23 | 1.1 | 2002 | 9 BC011365 | BC011365 Homo sapi |
| 33 | 23 | 1.1 | 2459 | 10 BC013482 | BC013482 Mus muscu |
| 34 | 23 | 1.1 | 2488 | 10 BC026628 | BC026628 Mus muscu |
| 35 | 23 | 1.1 | 2860 | 6 E12442 | E12442 DNR encodin |
| 36 | 23 | 1.1 | 29054 | 2 AC017237 | AC017237 Drosophill |
| 37 | 23 | 1.1 | 35470 | 2 AC116989 | AC116989 Dictyoste |
| 38 | 23 | 1.1 | 56342 | 2 AC103270 | AC103270 Rattus no |
| 39 | 23 | 1.1 | 63739 | 2 AC087621 | AC087621 Homo sapi |
| 40 | 23 | 1.1 | 69523 | 2 AC124305 | AC124305 Homo sapi |
| 41 | 23 | 1.1 | 74718 | 10 AC083858 | AC083858 Mus muscu |
| 42 | 23 | 1.1 | 83543 | 9 AC117427 | AC117427 Homo sapi |
| 43 | 23 | 1.1 | 92910 | 9 AL590368 | AL590368 Human DNA |
| 44 | 23 | 1.1 | 97847 | 9 HS874C20 | HS874C20 Human DNA |
| 45 | 23 | 1.1 | 109151 | 2 AC017637 | AC017637 Drosophill |
| 46 | 23 | 1.1 | 111689 | 2 AC128720 | AC128720 Rattus no |
| 47 | 23 | 1.1 | 133902 | 2 AC095664 | AC095664 Rattus no |
| 48 | 23 | 1.1 | 141017 | 2 AC116962 | AC116962 Dictyoste |
| 49 | 23 | 1.1 | 143351 | 2 AC099282 | AC099282 Rattus no |
| 50 | 23 | 1.1 | 159170 | 2 AC115914 | AC115914 Mus muscu |
| 51 | 23 | 1.1 | 160629 | 3 AC023715 | AC023715 Drosophill |
| 52 | 23 | 1.1 | 161278 | 3 AC008223 | AC008223 Drosophill |
| 53 | 23 | 1.1 | 164637 | 2 AC016192 | AC016192 Homo sapi |
| 54 | 23 | 1.1 | 172965 | 2 AC023367 | AC023367 Homo sapi |
| 55 | 23 | 1.1 | 172970 | 2 AC111112 | AC111112 Mus muscu |
| 56 | 23 | 1.1 | 177080 | 2 AC027288 | AC027288 Homo sapi |
| 57 | 23 | 1.1 | 177142 | 9 AC004582 | AC004582 Homo sapi |
| 58 | 23 | 1.1 | 177357 | 2 AC124078 | AC124078 Human Chr |
| 59 | 23 | 1.1 | 179618 | 2 AC068212 | AC068212 Homo sapi |
| 60 | 23 | 1.1 | 180169 | 9 AL590101 | AL590101 Human DNA |
| 61 | 23 | 1.1 | 180692 | 2 AC123320 | AC123320 Rattus no |
| 62 | 23 | 1.1 | 184551 | 2 AC117869 | AC117869 Rattus no |
| 63 | 23 | 1.1 | 185671 | 9 AC084193 | AC084193 Homo sapi |
| 64 | 23 | 1.1 | 187543 | 9 AC005678 | AC005678 Homo sapi |
| 65 | 23 | 1.1 | 190153 | 3 AC008201 | AC008201 Drosophill |

| | | | | | | | | | | | | | |
|---|-----|----|-----|--------|----|-----------|-------|----|-----|--------|----|------------|---------------------|
| C | 65 | 23 | 1.1 | 205429 | 2 | AC005506 | 139 | 22 | 1.1 | 94234 | 2 | AC122296 | AC122296 Rattus no |
| C | 67 | 23 | 1.1 | 215523 | 2 | AL844180 | C 140 | 22 | 1.1 | 96433 | 2 | DMB519 | AL121800 Drosophill |
| C | 68 | 23 | 1.1 | 217208 | 2 | AC121927 | C 141 | 22 | 1.1 | 97799 | 2 | AL354890 | AL354890 Homo sapi |
| C | 69 | 23 | 1.1 | 237950 | 2 | AC127315 | C 142 | 22 | 1.1 | 99758 | 2 | AC097051 | AC097051 Rattus no |
| C | 70 | 23 | 1.1 | 239327 | 3 | AE003744 | C 143 | 22 | 1.1 | 100726 | 2 | AC116961 | AC116961 Dicyoste |
| C | 71 | 23 | 1.1 | 288687 | 2 | AC125315 | C 144 | 22 | 1.1 | 102064 | 2 | AC002558 | AC002558 Homo sapi |
| C | 72 | 23 | 1.1 | 319551 | 3 | AE003432 | C 145 | 22 | 1.1 | 104014 | 2 | AC116921 | AC116921 Dicyoste |
| C | 73 | 22 | 1.1 | 3302 | 3 | DDACT15P | C 146 | 22 | 1.1 | 104483 | 2 | AC106977 | AC106977 Rattus no |
| C | 74 | 22 | 1.1 | 707 | 3 | AY075486 | C 147 | 22 | 1.1 | 107609 | 2 | AC118062 | AC118062 Homo sapi |
| C | 75 | 22 | 1.1 | 934 | 3 | DD025660 | C 148 | 22 | 1.1 | 110000 | 2 | PEMALALP1 | Continuation (2 of |
| C | 76 | 22 | 1.1 | 1435 | 3 | AF378371 | C 149 | 22 | 1.1 | 110000 | 10 | AE014177_1 | AE014177_1 |
| C | 77 | 22 | 1.1 | 1477 | 3 | AY089472 | C 150 | 22 | 1.1 | 110000 | 30 | AC078780_0 | AC078780 Homo sapi |
| C | 78 | 22 | 1.1 | 1629 | 5 | AF124511 | C 151 | 22 | 1.1 | 111104 | 9 | AC117389 | AC117389 Homo sapi |
| C | 79 | 22 | 1.1 | 1740 | 5 | HUMB51PRO | C 152 | 22 | 1.1 | 111173 | 3 | AL513204 | AL513204 Human DNA |
| C | 80 | 22 | 1.1 | 1782 | 6 | AX401786 | C 153 | 22 | 1.1 | 113880 | 9 | PEMAL3P4 | AL008970 Plasmodiu |
| C | 81 | 22 | 1.1 | 1782 | 10 | RNU32208 | C 154 | 22 | 1.1 | 114748 | 2 | AC103434 | AC103434 Rattus no |
| C | 82 | 22 | 1.1 | 1801 | 10 | BC012744 | C 155 | 22 | 1.1 | 114748 | 2 | AC103434 | AC103434 Rattus no |
| C | 83 | 22 | 1.1 | 1801 | 9 | AF289495 | C 156 | 22 | 1.1 | 115018 | 9 | AL162595 | AL162595 Human DNA |
| C | 84 | 22 | 1.1 | 2007 | 9 | AF289495 | C 157 | 22 | 1.1 | 115018 | 9 | AL162595 | AL162595 Human DNA |
| C | 85 | 22 | 1.1 | 2115 | 9 | AK024885 | C 158 | 22 | 1.1 | 115545 | 2 | AC119334 | AC119334 Rattus no |
| C | 86 | 22 | 1.1 | 2220 | 3 | DDACT15 | C 159 | 22 | 1.1 | 117283 | 2 | CNS01RIL | AL163533 Human chr |
| C | 87 | 22 | 1.1 | 2230 | 3 | DDCAMPI | C 160 | 22 | 1.1 | 119696 | 9 | HSDJ777L9 | AC1088240 Human DNA |
| C | 88 | 22 | 1.1 | 2499 | 10 | BC008135 | C 161 | 22 | 1.1 | 121539 | 2 | AC1088240 | AC1088240 Rattus no |
| C | 89 | 22 | 1.1 | 2558 | 3 | PEBDAS | C 162 | 22 | 1.1 | 123280 | 2 | AC117076 | AC117076 Dicyoste |
| C | 90 | 22 | 1.1 | 2610 | 9 | HSB80395 | C 163 | 22 | 1.1 | 126143 | 2 | AC098343 | AC098343 Rattus no |
| C | 91 | 22 | 1.1 | 2678 | 8 | AF360344 | C 164 | 22 | 1.1 | 131259 | 2 | AP001805 | AP001805 Homo sapi |
| C | 92 | 22 | 1.1 | 2691 | 9 | BC017210 | C 165 | 22 | 1.1 | 134290 | 2 | HSJ543J13 | AL121939 Human DNA |
| C | 93 | 22 | 1.1 | 2844 | 10 | BC027428 | C 166 | 22 | 1.1 | 137712 | 2 | AC102683 | AC102683 Mus muscu |
| C | 94 | 22 | 1.1 | 2941 | 9 | AK025562 | C 167 | 22 | 1.1 | 142232 | 8 | AC107315 | AC107315 Genomic s |
| C | 95 | 22 | 1.1 | 3128 | 6 | A21625 | C 168 | 22 | 1.1 | 143446 | 9 | AC079855 | AC079855 Homo sapi |
| C | 96 | 22 | 1.1 | 3357 | 6 | AX277591 | C 169 | 22 | 1.1 | 143100 | 2 | AC024641 | AC024641 Homo sapi |
| C | 97 | 22 | 1.1 | 3597 | 9 | HSB800936 | C 170 | 22 | 1.1 | 146399 | 2 | AC025835 | AC025835 Homo sapi |
| C | 98 | 22 | 1.1 | 4814 | 3 | AY075427 | C 171 | 22 | 1.1 | 146487 | 9 | AC010803 | AC010803 Homo sapi |
| C | 99 | 22 | 1.1 | 4830 | 3 | DME420377 | C 172 | 22 | 1.1 | 147192 | 2 | AC116925 | AC116925 Dicyoste |
| C | 100 | 22 | 1.1 | 6105 | 12 | ASPDAXH | C 173 | 22 | 1.1 | 147750 | 2 | AC110930 | AC110930 Rattus no |
| C | 101 | 22 | 1.1 | 6108 | 12 | ASPDAXH | C 174 | 22 | 1.1 | 148557 | 2 | AC116827 | AC116827 Mus muscu |
| C | 102 | 22 | 1.1 | 6113 | 12 | ASPDAXH | C 175 | 22 | 1.1 | 151365 | 2 | AC010382 | AC010382 Homo sapi |
| C | 103 | 22 | 1.1 | 6114 | 12 | ASPDAXH | C 176 | 22 | 1.1 | 151476 | 9 | AC116591 | AC116591 Mus muscu |
| C | 104 | 22 | 1.1 | 6146 | 12 | AF269236 | C 177 | 22 | 1.1 | 152409 | 2 | PEMALP1 | AL031744 Plasmodiu |
| C | 105 | 22 | 1.1 | 6872 | 12 | AF269235 | C 178 | 22 | 1.1 | 153360 | 2 | AC129073 | AC129073 Felis cat |
| C | 106 | 22 | 1.1 | 8574 | 3 | AF057019 | C 179 | 22 | 1.1 | 153360 | 2 | AC095290 | AC095290 Rattus no |
| C | 107 | 22 | 1.1 | 10452 | 12 | AF269237 | C 180 | 22 | 1.1 | 154371 | 9 | AP003466 | AP003466 Homo sapi |
| C | 108 | 22 | 1.1 | 1166 | 12 | AF269238 | C 181 | 22 | 1.1 | 155020 | 2 | AC106390 | AC106390 Rattus no |
| C | 109 | 22 | 1.1 | 12029 | 3 | AE001385 | C 182 | 22 | 1.1 | 155191 | 2 | AC119025 | AC119025 Rattus no |
| C | 110 | 22 | 1.1 | 12229 | 3 | AE001377 | C 183 | 22 | 1.1 | 155222 | 10 | AL683881 | AL683881 Mouse DNA |
| C | 111 | 22 | 1.1 | 12940 | 3 | AE001386 | C 184 | 22 | 1.1 | 156506 | 2 | AC004153 | AC004153 Plasmodiu |
| C | 112 | 22 | 1.1 | 17706 | 2 | AC014224 | C 185 | 22 | 1.1 | 156506 | 2 | AC068038 | AC068038 Homo sapi |
| C | 113 | 22 | 1.1 | 23676 | 2 | AC115599 | C 186 | 22 | 1.1 | 157069 | 2 | AC123209 | AC123209 Rattus no |
| C | 114 | 22 | 1.1 | 27314 | 9 | AC091224 | C 187 | 22 | 1.1 | 157102 | 2 | AC116409 | AC116409 Homo sapi |
| C | 115 | 22 | 1.1 | 36521 | 2 | AC053516 | C 188 | 22 | 1.1 | 157124 | 2 | AL355813 | AL355813 Human DNA |
| C | 116 | 22 | 1.1 | 37572 | 2 | AC020203 | C 189 | 22 | 1.1 | 157150 | 9 | AL355813 | AL355813 Homo sapi |
| C | 117 | 22 | 1.1 | 37572 | 2 | AC014052 | C 190 | 22 | 1.1 | 157336 | 2 | AC095253 | AC095253 Rattus no |
| C | 118 | 22 | 1.1 | 3765 | 2 | AC109977 | C 191 | 22 | 1.1 | 157361 | 2 | AC106572 | AC106572 Rattus no |
| C | 119 | 22 | 1.1 | 44700 | 9 | AL590825 | C 192 | 22 | 1.1 | 157976 | 2 | AC127476 | AC127476 Felis cat |
| C | 120 | 22 | 1.1 | 46879 | 2 | AC015095 | C 193 | 22 | 1.1 | 158750 | 2 | AC067981 | AC067981 Homo sapi |
| C | 121 | 22 | 1.1 | 53380 | 2 | AL135792 | C 194 | 22 | 1.1 | 158832 | 2 | AC097950 | AC097950 Rattus no |
| C | 122 | 22 | 1.1 | 54527 | 2 | AC015174 | C 195 | 22 | 1.1 | 159774 | 2 | AC099367 | AC099367 Rattus no |
| C | 123 | 22 | 1.1 | 55038 | 2 | AC015270 | C 196 | 22 | 1.1 | 160001 | 3 | AC113619 | AC113619 Drosophill |
| C | 124 | 22 | 1.1 | 55470 | 2 | AC116968 | C 197 | 22 | 1.1 | 160330 | 2 | AC025631 | AC025631 Homo sapi |
| C | 125 | 22 | 1.1 | 59866 | 2 | AC100572 | C 198 | 22 | 1.1 | 160759 | 2 | AC117082 | AC117082 Dicyoste |
| C | 126 | 22 | 1.1 | 65201 | 3 | AC004059 | C 199 | 22 | 1.1 | 161201 | 2 | AC025673 | AC025673 Homo sapi |
| C | 127 | 22 | 1.1 | 67297 | 3 | AC005135 | C 200 | 22 | 1.1 | 161437 | 2 | AC093434 | AC093434 Homo sapi |
| C | 128 | 22 | 1.1 | 68464 | 9 | AC005248 | C 201 | 22 | 1.1 | 163035 | 2 | AC006405 | AC006405 Homo sapi |
| C | 129 | 22 | 1.1 | 69480 | 9 | AC092587 | C 202 | 22 | 1.1 | 163457 | 2 | AC111523 | AC111523 Rattus no |
| C | 130 | 22 | 1.1 | 69511 | 2 | AC094686 | C 203 | 22 | 1.1 | 163879 | 3 | AC099310 | AC099310 Drosophill |
| C | 131 | 22 | 1.1 | 70578 | 2 | AC100228 | C 204 | 22 | 1.1 | 164349 | 2 | AC123886 | AC123886 Homo sapi |
| C | 132 | 22 | 1.1 | 73763 | 2 | AC111965 | C 205 | 22 | 1.1 | 165732 | 2 | AP004388 | AP004388 Homo sapi |
| C | 133 | 22 | 1.1 | 77601 | 9 | HS45P1 | C 206 | 22 | 1.1 | 165929 | 2 | AL844157 | AL844157 Mus muscu |
| C | 134 | 22 | 1.1 | 82469 | 8 | AC116330 | C 207 | 22 | 1.1 | 167944 | 2 | AC079805 | AC079805 Homo sapi |
| C | 135 | 22 | 1.1 | 83865 | 8 | AB006700 | C 208 | 22 | 1.1 | 168185 | 2 | AC124357 | AC124357 Mus muscu |
| C | 136 | 22 | 1.1 | 91355 | 2 | AL391417 | C 209 | 22 | 1.1 | 168426 | 2 | AL731820 | AL731820 Mouse DNA |
| C | 137 | 22 | 1.1 | 93897 | 2 | AC129831 | C 210 | 22 | 1.1 | 168773 | 10 | | |
| C | 138 | 22 | 1.1 | 93913 | 2 | AC122576 | C 211 | 22 | 1.1 | 168773 | 10 | | |

| | | | | | | | | | | | | | |
|-------|----|-----|--------|----|-----------|-----------|-------|-----|---------|----|-----------|-----------|------------|
| C 212 | 22 | 1.1 | 169161 | 2 | AC120589 | Rattus no | 285 | 1.1 | 209483 | 9 | CNS05TDV | AL357153 | Human chr |
| C 213 | 22 | 1.1 | 169190 | 2 | AC114776 | Homo sapi | 286 | 1.1 | 209636 | 2 | AL732412 | AL732412 | Mus muscu |
| C 214 | 22 | 1.1 | 169308 | 2 | AC124700 | Mus muscu | C 287 | 1.1 | 210336 | 2 | AC022171 | AC022171 | Homo sapi |
| C 215 | 22 | 1.1 | 169469 | 2 | AC037195 | Homo sapi | 288 | 1.1 | 210631 | 2 | AC129184 | AC129184 | Mus muscu |
| C 216 | 22 | 1.1 | 169510 | 2 | AC129168 | Rattus no | 289 | 1.1 | 2113045 | 2 | AC093640 | AC093640 | Homo sapi |
| C 217 | 22 | 1.1 | 169774 | 2 | AC0129622 | Rattus no | 290 | 1.1 | 216347 | 2 | AL663095 | AL663095 | Mus muscu |
| C 218 | 22 | 1.1 | 169794 | 2 | AC004688 | Plasmodiu | 291 | 1.1 | 216408 | 2 | AC092466 | AC092466 | Homo sapi |
| C 219 | 22 | 1.1 | 171020 | 2 | AC036225 | Homo sapi | C 292 | 1.1 | 217164 | 2 | AL844482 | AL844482 | Mus muscu |
| C 220 | 22 | 1.1 | 171286 | 3 | AC008226 | Drosophi1 | C 293 | 1.1 | 217353 | 2 | AC123033 | AC123033 | Mus muscu |
| C 221 | 22 | 1.1 | 171390 | 9 | AC027315 | Homo sapi | 294 | 1.1 | 217409 | 9 | AC123901 | AC123901 | Homo sapi |
| C 222 | 22 | 1.1 | 171471 | 2 | AC113073 | Mus muscu | C 295 | 1.1 | 217743 | 9 | AC116293 | AC116293 | Rattus no |
| C 223 | 22 | 1.1 | 171873 | 2 | AC024995 | Homo sapi | 296 | 1.1 | 222605 | 2 | AC010973 | AC010973 | Homo sapi |
| C 224 | 22 | 1.1 | 172676 | 2 | AC087736 | Pan trogl | 297 | 1.1 | 223561 | 2 | AC121574 | AC121574 | Mus muscu |
| C 225 | 22 | 1.1 | 172714 | 2 | AL845489 | Mus muscu | C 298 | 1.1 | 224879 | 9 | AC008735 | AC008735 | Homo sapi |
| C 226 | 22 | 1.1 | 173188 | 3 | AC093196 | Drosophi1 | C 299 | 1.1 | 224448 | 2 | PFMAL14P4 | PFMAL14P4 | AC027284 |
| C 227 | 22 | 1.1 | 174323 | 2 | AC092754 | Homo sapi | 300 | 1.1 | 224840 | 2 | AC027284 | AC027284 | Mus muscu |
| C 228 | 22 | 1.1 | 174378 | 9 | AC020983 | Homo sapi | 301 | 1.1 | 226013 | 2 | AC073680 | AC073680 | Mus muscu |
| C 229 | 22 | 1.1 | 174579 | 2 | AC016666 | Homo sapi | C 302 | 1.1 | 226013 | 2 | AC073680 | AC073680 | Mus muscu |
| C 230 | 22 | 1.1 | 174594 | 2 | AC098986 | Rattus no | 303 | 1.1 | 227506 | 9 | AC079969 | AC079969 | Homo sapi |
| C 231 | 22 | 1.1 | 175856 | 2 | AC102209 | Mus muscu | 304 | 1.1 | 229045 | 5 | AC008680 | AC008680 | Homo sapi |
| C 232 | 22 | 1.1 | 176095 | 2 | AC019346 | Homo sapi | 305 | 1.1 | 230110 | 2 | AC122396 | AC122396 | Mus muscu |
| C 233 | 22 | 1.1 | 177203 | 2 | AC055784 | Homo sapi | C 306 | 1.1 | 233573 | 2 | AC091340 | AC091340 | Rattus no |
| C 234 | 22 | 1.1 | 178626 | 2 | AC110511 | Mus muscu | C 307 | 1.1 | 234634 | 3 | AE003740 | AE003740 | Drosophi1 |
| C 235 | 22 | 1.1 | 179039 | 2 | AC102295 | Mus muscu | 308 | 1.1 | 236246 | 2 | AC099068 | AC099068 | Rattus no |
| C 236 | 22 | 1.1 | 179396 | 3 | AC022351 | Drosophi1 | C 309 | 1.1 | 238534 | 2 | AC020965 | AC020965 | Mus muscu |
| C 237 | 22 | 1.1 | 179623 | 9 | AC013429 | Homo sapi | 310 | 1.1 | 245802 | 2 | AC006279 | AC006279 | Plasmodiu |
| C 238 | 22 | 1.1 | 179795 | 3 | AC105351 | Drosophi1 | C 311 | 1.1 | 258661 | 2 | AC129078 | AC129078 | Mus muscu |
| C 239 | 22 | 1.1 | 179902 | 2 | AP003630 | Oryza sat | C 312 | 1.1 | 265236 | 3 | AE003626 | AE003626 | Drosophi1 |
| C 240 | 22 | 1.1 | 180790 | 2 | AC106172 | Rattus no | C 313 | 1.1 | 268147 | 2 | AC116966 | AC116966 | Drosophi1 |
| C 241 | 22 | 1.1 | 181039 | 2 | AC106915 | Rattus no | C 314 | 1.1 | 268499 | 2 | AC098250 | AC098250 | Rattus no |
| C 242 | 22 | 1.1 | 181250 | 9 | AC090403 | Homo sapi | 315 | 1.1 | 292420 | 2 | AC020873 | AC020873 | Mus muscu |
| C 243 | 22 | 1.1 | 181328 | 2 | AC104945 | Homo sapi | C 316 | 1.1 | 293431 | 2 | PFMAL13P4 | PFMAL13P4 | AE003502 |
| C 244 | 22 | 1.1 | 182366 | 2 | AC041041 | Homo sapi | 317 | 1.1 | 301980 | 3 | AE003502 | AE003502 | Drosophi1 |
| C 245 | 22 | 1.1 | 182600 | 2 | AC034126 | Homo sapi | C 318 | 1.1 | 306135 | 3 | AE003436 | AE003436 | Drosophi1 |
| C 246 | 22 | 1.1 | 182622 | 2 | AC090567 | Homo sapi | 319 | 1.1 | 309023 | 5 | AE003671 | AE003671 | Drosophi1 |
| C 247 | 22 | 1.1 | 182660 | 2 | AC012416 | Homo sapi | 320 | 1.1 | 310779 | 2 | AC005140 | AC005140 | Plasmodiu |
| C 248 | 22 | 1.1 | 183982 | 2 | AC009825 | Homo sapi | 321 | 1.1 | 310909 | 2 | AE003427 | AE003427 | Drosophi1 |
| C 249 | 22 | 1.1 | 184026 | 2 | AC022176 | Homo sapi | 322 | 1.1 | 316084 | 2 | AC127345 | AC127345 | Mus muscu |
| C 250 | 22 | 1.1 | 185937 | 2 | AC094540 | Rattus no | 323 | 1.1 | 321003 | 2 | PFMAL1P3 | PFMAL1P3 | AE003476 |
| C 251 | 22 | 1.1 | 186092 | 9 | AC020556 | Homo sapi | C 324 | 1.1 | 195 | 9 | HS79D1R | HS79D1R | AE003476 |
| C 252 | 22 | 1.1 | 186323 | 9 | AC009405 | Homo sapi | C 325 | 1.1 | 196 | 9 | HS47F4R | HS47F4R | AE003476 |
| C 253 | 22 | 1.1 | 186559 | 10 | AL606742 | Mouse DNA | 326 | 1.1 | 196 | 9 | HS79D1R | HS79D1R | AE003476 |
| C 254 | 22 | 1.1 | 187730 | 2 | AC110613 | Homo sapi | 327 | 1.1 | 294 | 3 | AF082830 | AF082830 | Taenia so |
| C 255 | 22 | 1.1 | 188246 | 9 | AC013565 | Homo sapi | 328 | 1.1 | 284 | 6 | AX083370 | AX083370 | Sequence |
| C 256 | 22 | 1.1 | 188459 | 3 | AC008312 | Drosophi1 | 329 | 1.1 | 329 | 3 | AF082829 | AF082829 | Taenia so |
| C 257 | 22 | 1.1 | 188835 | 2 | AC104153 | Papio cyn | 330 | 1.1 | 329 | 3 | AF257776 | AF257776 | Taenia so |
| C 258 | 22 | 1.1 | 190008 | 2 | AC125917 | Rattus no | 331 | 1.1 | 364 | 3 | AF163972 | AF163972 | Taenia so |
| C 259 | 22 | 1.1 | 190727 | 9 | AC068724 | Homo sapi | 332 | 1.1 | 372 | 3 | AE216695 | AE216695 | Taenia so |
| C 260 | 22 | 1.1 | 190911 | 2 | AC095406 | Rattus no | C 333 | 1.1 | 402 | 8 | AE427562 | AE427562 | Bikinia a |
| C 261 | 22 | 1.1 | 191093 | 10 | AL671765 | Mouse DNA | C 334 | 1.1 | 410 | 6 | AF427568 | AF427568 | Bikinia g |
| C 262 | 22 | 1.1 | 191312 | 2 | AP001030 | Homo sapi | C 335 | 1.1 | 411 | 6 | AX330823 | AX330823 | Sequence |
| C 263 | 22 | 1.1 | 191750 | 2 | AC122498 | Mus muscu | C 336 | 1.1 | 411 | 6 | AX408781 | AX408781 | Sequence |
| C 264 | 22 | 1.1 | 193038 | 2 | AC128823 | Rattus no | 337 | 1.1 | 415 | 3 | AB044083 | AB044083 | Taenia so |
| C 265 | 22 | 1.1 | 193674 | 2 | AC125111 | Mus muscu | C 338 | 1.1 | 437 | 8 | AF427592 | AF427592 | Tetrahel1 |
| C 266 | 22 | 1.1 | 194000 | 2 | AC000016 | Homo sapi | C 339 | 1.1 | 439 | 8 | AF427578 | AF427578 | Aphanocai |
| C 267 | 22 | 1.1 | 194286 | 2 | AC129160 | Rattus no | 340 | 1.1 | 439 | 8 | BC031337 | BC031337 | Homo sapi |
| C 268 | 22 | 1.1 | 194551 | 2 | AC124924 | Rattus no | C 341 | 1.1 | 455 | 8 | AF427581 | AF427581 | Homo sapi |
| C 269 | 22 | 1.1 | 194552 | 2 | AC108327 | Rattus no | C 342 | 1.1 | 456 | 8 | AF427548 | AF427548 | Brechys-te |
| C 270 | 22 | 1.1 | 194633 | 2 | AC090362 | Homo sapi | C 343 | 1.1 | 457 | 8 | AF427363 | AF427363 | Bikinia d |
| C 271 | 22 | 1.1 | 195460 | 2 | AL672289 | Homo sapi | C 344 | 1.1 | 458 | 8 | AF427557 | AF427557 | Juldeinat |
| C 272 | 22 | 1.1 | 195727 | 2 | AC036234 | Homo sapi | C 345 | 1.1 | 458 | 8 | AF427556 | AF427556 | Bikinia d |
| C 273 | 22 | 1.1 | 195744 | 2 | AC092002 | Homo sapi | C 346 | 1.1 | 465 | 8 | AF427575 | AF427575 | Bikinia m |
| C 274 | 22 | 1.1 | 196000 | 2 | AL354749 | Human DNA | C 347 | 1.1 | 465 | 8 | AF427580 | AF427580 | Aphanocai |
| C 275 | 22 | 1.1 | 198124 | 2 | AC111459 | Rattus no | C 348 | 1.1 | 466 | 8 | AF427544 | AF427544 | Aphanocai |
| C 276 | 22 | 1.1 | 198487 | 2 | AC128485 | Rattus no | C 349 | 1.1 | 467 | 8 | AF427579 | AF427579 | Aphanocai |
| C 277 | 22 | 1.1 | 199551 | 2 | AC006281 | Plasmodiu | C 350 | 1.1 | 477 | 8 | AF427577 | AF427577 | Aphanocai |
| C 278 | 22 | 1.1 | 202904 | 2 | AP002077 | Homo sapi | C 351 | 1.1 | 484 | 8 | AF427559 | AF427559 | Aphanocai |
| C 279 | 22 | 1.1 | 203493 | 9 | AC058803 | Homo sapi | 352 | 1.1 | 647 | 8 | AF370180 | AF370180 | Arabiidops |
| C 280 | 22 | 1.1 | 203924 | 2 | AC012159 | Drosophi1 | 353 | 1.1 | 660 | 9 | BC020826 | BC020826 | Homo sapi |
| C 281 | 22 | 1.1 | 204652 | 2 | PFMAL13P6 | AC005506 | 354 | 1.1 | 663 | 10 | AY113211 | AY113211 | Drosophi1 |
| C 282 | 22 | 1.1 | 205429 | 2 | AC005506 | Plasmodiu | 355 | 1.1 | 764 | 3 | BC002095 | BC002095 | Mus muscu |
| C 283 | 22 | 1.1 | 206055 | 2 | AL844208 | Mus muscu | 356 | 1.1 | 808 | 10 | AF471666 | AF471666 | illex main |
| C 284 | 22 | 1.1 | 208671 | 2 | AC126498 | Rattus no | C 357 | 1.1 | 853 | 8 | AF471666 | AF471666 | illex main |

| | | | | | | | | | | | | | |
|-----|----|-----|-------|----|------------|---------------------|-------|----|-----|--------|---|------------|---------------------|
| 358 | 21 | 1.0 | 873 | 11 | CNS06E2J | AL395873 T3 end of | c 431 | 21 | 1.0 | 93735 | 8 | ATAC008261 | AC008261 Arabidops |
| 359 | 21 | 1.0 | 928 | 3 | BC029527 | BC029527 Homo sapi | c 432 | 21 | 1.0 | 95560 | 2 | AC103136 | AC103136 Rattus no |
| 360 | 21 | 1.0 | 1042 | 3 | AY070627 | AY070627 Drosophill | c 433 | 21 | 1.0 | 98062 | 2 | RN516010 | AL003713 Rattus no |
| 361 | 21 | 1.0 | 1101 | 10 | BC022557 | BC022557 Mus muscu | c 434 | 21 | 1.0 | 99725 | 8 | ATF18674 | AL021637 Arabidops |
| 362 | 21 | 1.0 | 1165 | 8 | AY065280 | AY065280 Arabidops | c 435 | 21 | 1.0 | 99920 | 2 | AC096218 | AC096218 Rattus no |
| 363 | 21 | 1.0 | 1280 | 3 | AY061225 | AY061225 Drosophill | c 436 | 21 | 1.0 | 100409 | 2 | AC1033280 | AC1033280 Rattus no |
| 364 | 21 | 1.0 | 1282 | 10 | MM0238878 | AJ238978 Mus muscu | c 437 | 21 | 1.0 | 100511 | 2 | AC010774 | AC010774 Homo sapi |
| 365 | 21 | 1.0 | 1364 | 9 | BC030144 | BC030144 Homo sapi | c 438 | 21 | 1.0 | 100685 | 8 | AC023754 | AC023754 Arabidops |
| 366 | 21 | 1.0 | 1439 | 10 | BC005457 | BC005457 Mus muscu | c 439 | 21 | 1.0 | 102241 | 2 | AC097198 | AC097198 Rattus no |
| 367 | 21 | 1.0 | 1449 | 8 | TAFSBDCL | X07742 Wheat chlor | c 440 | 21 | 1.0 | 104001 | 8 | AC023528 | AC023528 Rattus no |
| 368 | 21 | 1.0 | 1469 | 6 | AX375585 | AX375585 Sequence | c 441 | 21 | 1.0 | 104163 | 8 | AC010155 | AC010155 Genomic s |
| 369 | 21 | 1.0 | 1486 | 6 | BC007016 | BC007016 Homo sapi | c 442 | 21 | 1.0 | 105507 | 2 | AC109113 | AC109113 Rattus no |
| 370 | 21 | 1.0 | 1547 | 8 | AF160729 | AF160729 Arabidops | c 443 | 21 | 1.0 | 106490 | 2 | AC095569 | AC095569 Rattus no |
| 371 | 21 | 1.0 | 1593 | 8 | BC008841 | BC008841 Homo sapi | c 444 | 21 | 1.0 | 110000 | 2 | AC095560-1 | Continuation (2 of |
| 372 | 21 | 1.0 | 1594 | 8 | CEY17297 | Y17297 Cunninghame | c 445 | 21 | 1.0 | 111682 | 2 | AC005162 | AC005162 Homo sapi |
| 373 | 21 | 1.0 | 1746 | 8 | BC011614 | BC011614 Homo sapi | c 446 | 21 | 1.0 | 115028 | 2 | AC099177 | AC099177 Rattus no |
| 374 | 21 | 1.0 | 1793 | 8 | AY070432 | AY070432 Arabidops | c 447 | 21 | 1.0 | 118426 | 2 | HS431C21 | AC099177 Rattus no |
| 375 | 21 | 1.0 | 1967 | 3 | AY060688 | AY060688 Drosophill | c 448 | 21 | 1.0 | 119035 | 2 | AC096954 | AC096954 Rattus no |
| 376 | 21 | 1.0 | 2041 | 3 | ALRDNA1 | X05836 Ascaris lum | c 449 | 21 | 1.0 | 119063 | 2 | AC027311 | AC027311 Homo sapi |
| 377 | 21 | 1.0 | 2145 | 3 | AY070598 | AY070598 Drosophill | c 450 | 21 | 1.0 | 123039 | 2 | AC116268 | AC116268 Rattus no |
| 378 | 21 | 1.0 | 2146 | 3 | AY113401 | AY113401 Drosophill | c 451 | 21 | 1.0 | 125144 | 2 | AC099359 | AC099359 Rattus no |
| 379 | 21 | 1.0 | 2151 | 10 | BC005571 | BC005571 Mus muscu | c 452 | 21 | 1.0 | 129218 | 2 | AC0933280 | AC0933280 Homo sapi |
| 380 | 21 | 1.0 | 2230 | 8 | CFP450 | AF236871 Mus muscu | c 453 | 21 | 1.0 | 129719 | 9 | AL358113 | AL358113 Human DNA |
| 381 | 21 | 1.0 | 2262 | 3 | AF079448 | X76226 Candida mal | c 454 | 21 | 1.0 | 130834 | 2 | AC115231 | AC115231 Rattus no |
| 382 | 21 | 1.0 | 2403 | 5 | AF014370 | AF014370 Dactyoste | c 455 | 21 | 1.0 | 130923 | 2 | AC116295 | AC116295 Rattus no |
| 383 | 21 | 1.0 | 2423 | 5 | AF014370 | AF014370 Danio rer | c 456 | 21 | 1.0 | 131245 | 2 | AC113320 | AC113320 Human DNA |
| 384 | 21 | 1.0 | 2823 | 9 | BC009981 | BC009981 Homo sapi | c 457 | 21 | 1.0 | 132596 | 2 | AC013046 | AC013046 Drosophill |
| 385 | 21 | 1.0 | 2912 | 9 | AF261655 | AF261655 Homo sapi | c 458 | 21 | 1.0 | 132966 | 2 | AC125403 | AC125403 Mus muscu |
| 386 | 21 | 1.0 | 3160 | 10 | AF102887 | AF102887 Mus muscu | c 459 | 21 | 1.0 | 133255 | 2 | AC121410 | AC121410 Rattus no |
| 387 | 21 | 1.0 | 3379 | 3 | AF111943 | AF111943 Dactyoste | c 460 | 21 | 1.0 | 133376 | 2 | AC087147 | AC087147 Mus muscu |
| 388 | 21 | 1.0 | 3391 | 10 | F32889S11 | AF328994 Mus muscu | c 461 | 21 | 1.0 | 133501 | 2 | AC116556 | AC116556 Dactyoste |
| 389 | 21 | 1.0 | 3479 | 3 | NEMRGA | M74584 Ascaris lum | c 462 | 21 | 1.0 | 134190 | 2 | AC128770 | AC128770 Rattus no |
| 390 | 21 | 1.0 | 3597 | 3 | AY118761 | AY118761 Drosophill | c 463 | 21 | 1.0 | 134372 | 2 | AC118807 | AC118807 Rattus no |
| 391 | 21 | 1.0 | 4110 | 3 | PFASERRP | M55428 P.falciparu | c 464 | 21 | 1.0 | 134545 | 8 | AB042240 | AB042240 Trilicium |
| 392 | 21 | 1.0 | 4155 | 8 | CMCHMGCOAR | L10390 Camptotheca | c 465 | 21 | 1.0 | 137154 | 2 | AL513082 | AL513082 Human DNA |
| 393 | 21 | 1.0 | 4226 | 10 | BC019138 | U20661 Dactyoste | c 466 | 21 | 1.0 | 138019 | 2 | AC121737 | AC121737 Rattus no |
| 394 | 21 | 1.0 | 4284 | 3 | DDU20651 | AL831998 Homo sapi | c 467 | 21 | 1.0 | 138859 | 2 | AL359076 | AL359076 Human DNA |
| 395 | 21 | 1.0 | 4638 | 9 | HSN803305 | AL831998 Homo sapi | c 468 | 21 | 1.0 | 139390 | 2 | AC110432 | AC110432 Rattus no |
| 396 | 21 | 1.0 | 4669 | 10 | AF029240 | AF029240 Rattus no | c 469 | 21 | 1.0 | 139555 | 2 | AC025636 | AC025636 Homo sapi |
| 397 | 21 | 1.0 | 5310 | 2 | AC012823 | AC012823 Drosophill | c 470 | 21 | 1.0 | 140714 | 2 | RN574E16 | AL003712 Rattus no |
| 398 | 21 | 1.0 | 6608 | 6 | AX346936 | AX346936 Sequence | c 471 | 21 | 1.0 | 143244 | 9 | AC091997 | AC091997 Homo sapi |
| 399 | 21 | 1.0 | 9098 | 3 | AF465306 | AF465306 Dactyoste | c 472 | 21 | 1.0 | 144476 | 2 | AC117283 | AC117283 Rattus no |
| 400 | 21 | 1.0 | 9671 | 1 | AE006059 | AE006059 Pasteurel | c 473 | 21 | 1.0 | 144542 | 2 | AC015819 | AC015819 Homo sapi |
| 401 | 21 | 1.0 | 12029 | 3 | AE001427 | AE001427 Dactyoste | c 474 | 21 | 1.0 | 145124 | 2 | AC022894 | AC022894 Homo sapi |
| 402 | 21 | 1.0 | 12355 | 3 | AC115605 | AC115605 Dactyoste | c 475 | 21 | 1.0 | 145577 | 2 | AC114233 | AC114233 Rattus no |
| 403 | 21 | 1.0 | 12593 | 3 | AE001418 | AE001418 Plasmodiu | c 476 | 21 | 1.0 | 147775 | 9 | AC015656 | AC015656 Homo sapi |
| 404 | 21 | 1.0 | 12919 | 3 | AC115580 | AC115580 Dactyoste | c 477 | 21 | 1.0 | 147882 | 2 | AC118083 | AC118083 Rattus no |
| 405 | 21 | 1.0 | 14760 | 3 | AE001410 | AE001410 Plasmodiu | c 478 | 21 | 1.0 | 149599 | 2 | AC112375 | AC112375 Rattus no |
| 406 | 21 | 1.0 | 15619 | 2 | AL590650 | AL590650 Human DNA | c 479 | 21 | 1.0 | 151455 | 2 | AC095571 | AC095571 Rattus no |
| 407 | 21 | 1.0 | 18010 | 2 | AC115607 | AC115607 Dactyoste | c 480 | 21 | 1.0 | 152116 | 2 | AC125089 | AC125089 Mus muscu |
| 408 | 21 | 1.0 | 20043 | 2 | AC116986 | AC116986 Dactyoste | c 481 | 21 | 1.0 | 152665 | 2 | AC027005 | AC027005 Homo sapi |
| 409 | 21 | 1.0 | 33404 | 2 | AC115684 | AC115684 Dactyoste | c 482 | 21 | 1.0 | 152717 | 2 | AC119505 | AC119505 Rattus no |
| 410 | 21 | 1.0 | 38872 | 2 | AC068113 | AC068113 Homo sapi | c 483 | 21 | 1.0 | 153004 | 9 | AC096754 | AC096754 Homo sapi |
| 411 | 21 | 1.0 | 43324 | 2 | AC034287 | AC034287 Mus muscu | c 484 | 21 | 1.0 | 153798 | 2 | AC108343 | AC108343 Rattus no |
| 412 | 21 | 1.0 | 48699 | 2 | AC115677 | AC115677 Dactyoste | c 485 | 21 | 1.0 | 153993 | 2 | AC012262 | AC012262 Homo sapi |
| 413 | 21 | 1.0 | 56099 | 2 | AC115598 | AC115598 Dactyoste | c 486 | 21 | 1.0 | 154076 | 2 | AC023273 | AC023273 Homo sapi |
| 414 | 21 | 1.0 | 58879 | 2 | AC090350 | AC090350 Homo sapi | c 487 | 21 | 1.0 | 154084 | 9 | AC092034 | AC092034 Homo sapi |
| 415 | 21 | 1.0 | 63117 | 2 | AC116505 | AC116505 Mus muscu | c 488 | 21 | 1.0 | 154222 | 2 | AC130631 | AC130631 Rattus no |
| 416 | 21 | 1.0 | 67757 | 9 | AL752388 | AL752388 Human DNA | c 489 | 21 | 1.0 | 154329 | 2 | DMR825B3 | DMR825B3 Drosophill |
| 417 | 21 | 1.0 | 69578 | 2 | AC118025 | AC118025 Mus muscu | c 490 | 21 | 1.0 | 154348 | 9 | AL691514 | AL691514 Human DNA |
| 418 | 21 | 1.0 | 71319 | 2 | AC009670 | AC009670 Homo sapi | c 491 | 21 | 1.0 | 154383 | 2 | AC113760 | AC113760 Rattus no |
| 419 | 21 | 1.0 | 72541 | 9 | AC007022 | AC007022 Homo sapi | c 492 | 21 | 1.0 | 155584 | 2 | AL161936 | AL161936 Human DNA |
| 420 | 21 | 1.0 | 75852 | 9 | AC074133 | AC074133 Homo sapi | c 493 | 21 | 1.0 | 155606 | 2 | AC105664 | AC105664 Rattus no |
| 421 | 21 | 1.0 | 76578 | 2 | AC109578 | AC109578 Rattus no | c 494 | 21 | 1.0 | 155690 | 2 | AC096042 | AC096042 Rattus no |
| 422 | 21 | 1.0 | 81766 | 2 | AC112134 | AC112134 Homo sapi | c 495 | 21 | 1.0 | 156358 | 2 | AC012298 | AC012298 Homo sapi |
| 423 | 21 | 1.0 | 82893 | 8 | AB022218 | AB022218 Arabidops | c 496 | 21 | 1.0 | 156359 | 2 | AC130743 | AC130743 Rattus no |
| 424 | 21 | 1.0 | 82893 | 8 | AC024081 | AC024081 Arabidops | c 497 | 21 | 1.0 | 157081 | 2 | AC023084 | AC023084 Homo sapi |
| 425 | 21 | 1.0 | 84194 | 8 | AB006701 | AB006701 Arabidops | c 498 | 21 | 1.0 | 157097 | 2 | AC024295 | AC024295 Homo sapi |
| 426 | 21 | 1.0 | 84648 | 2 | AC016461 | AC016461 Homo sapi | c 499 | 21 | 1.0 | 157674 | 2 | AP002470 | AP002470 Homo sapi |
| 427 | 21 | 1.0 | 86106 | 8 | AC011282 | AC011282 Homo sapi | c 500 | 21 | 1.0 | 158055 | 2 | AC127889 | AC127889 Rattus no |
| 428 | 21 | 1.0 | 88306 | 8 | NCRB7K22 | AL670543 Neurospor | c 501 | 21 | 1.0 | 158254 | 2 | AC112550 | AC112550 Rattus no |
| 429 | 21 | 1.0 | 92906 | 2 | AC096459 | AC096459 Rattus no | c 502 | 21 | 1.0 | 158558 | 2 | AC094305 | AC094305 Rattus no |
| 430 | 21 | 1.0 | 93581 | 9 | HS1059H15 | AL022100 Human DNA | c 503 | 21 | 1.0 | 158615 | 2 | AC117835 | AC117835 Rattus no |

| | | | | | | | | | | | | | | | | | |
|-----|---|-----|----|-----|--------|----|----------|----------|------------|-----|----|-----|--------|----|------------|------------|------------|
| 504 | C | 504 | 21 | 1.0 | 159505 | 2 | AC023142 | AC023142 | Homo sapi | 577 | 21 | 1.0 | 18981 | 2 | AC061986 | AC061986 | Homo sapi |
| 505 | C | 505 | 21 | 1.0 | 160673 | 10 | AL713977 | AL713977 | Mouse DNA | 578 | 21 | 1.0 | 190379 | 2 | AC062386 | AC062386 | Homo sapi |
| 506 | C | 506 | 21 | 1.0 | 161593 | 2 | AC114387 | AC114387 | Rattus no | 579 | 21 | 1.0 | 190511 | 2 | AC118096 | AC118096 | Rattus no |
| 507 | C | 507 | 21 | 1.0 | 161813 | 9 | CNS01RIG | AL163394 | Human chr | 580 | 21 | 1.0 | 190531 | 2 | AC124574 | AC124574 | Mus muscu |
| 508 | C | 508 | 21 | 1.0 | 163808 | 9 | AC015961 | AC015961 | Homo sapi | 581 | 21 | 1.0 | 190748 | 2 | AC120812 | AC120812 | Rattus no |
| 509 | C | 509 | 21 | 1.0 | 164007 | 9 | AC095155 | AC095155 | Rattus no | 582 | 21 | 1.0 | 191161 | 2 | AC107539 | AC107539 | Rattus no |
| 510 | C | 510 | 21 | 1.0 | 164018 | 9 | AL354868 | AL354868 | Human DNA | 583 | 21 | 1.0 | 191191 | 2 | HSAC000119 | HSAC000119 | Human BAC |
| 511 | C | 511 | 21 | 1.0 | 164423 | 9 | AC026828 | AC026828 | Homo sapi | 584 | 21 | 1.0 | 191342 | 3 | AC104140 | AC104140 | Drosophill |
| 512 | C | 512 | 21 | 1.0 | 164819 | 9 | AC092335 | AC092335 | Homo sapi | 585 | 21 | 1.0 | 191354 | 2 | AC102648 | AC102648 | Mus muscu |
| 513 | C | 513 | 21 | 1.0 | 164970 | 2 | AC098136 | AC098136 | Rattus no | 586 | 21 | 1.0 | 191563 | 2 | AC094883 | AC094883 | Rattus no |
| 514 | C | 514 | 21 | 1.0 | 165416 | 9 | AC034103 | AC034103 | Homo sapi | 587 | 21 | 1.0 | 192003 | 2 | AC084292 | AC084292 | Mus muscu |
| 515 | C | 515 | 21 | 1.0 | 165970 | 9 | AC098052 | AC098052 | Rattus no | 588 | 21 | 1.0 | 192087 | 2 | AC079335 | AC079335 | Homo sapi |
| 516 | C | 516 | 21 | 1.0 | 166341 | 2 | AL513486 | AL513486 | Homo sapi | 589 | 21 | 1.0 | 192138 | 2 | AC109381 | AC109381 | Rattus no |
| 517 | C | 517 | 21 | 1.0 | 166503 | 9 | AC023577 | AC023577 | Homo sapi | 590 | 21 | 1.0 | 192367 | 2 | AC107393 | AC107393 | Homo sapi |
| 518 | C | 518 | 21 | 1.0 | 166534 | 2 | AC034209 | AC034209 | Homo sapi | 591 | 21 | 1.0 | 192517 | 2 | AC094413 | AC094413 | Rattus no |
| 519 | C | 519 | 21 | 1.0 | 166692 | 2 | AC023475 | AC023475 | Homo sapi | 592 | 21 | 1.0 | 192959 | 2 | AC117065 | AC117065 | Rattus no |
| 520 | C | 520 | 21 | 1.0 | 166839 | 2 | AC130987 | AC130987 | Rattus no | 593 | 21 | 1.0 | 193118 | 2 | AC067728 | AC067728 | Homo sapi |
| 521 | C | 521 | 21 | 1.0 | 167694 | 9 | AC113137 | AC113137 | Homo sapi | 594 | 21 | 1.0 | 193469 | 2 | AC097966 | AC097966 | Homo sapi |
| 522 | C | 522 | 21 | 1.0 | 168154 | 2 | AC014098 | AC014098 | Rattus no | 595 | 21 | 1.0 | 193659 | 10 | AC122438 | AC122438 | Mus muscu |
| 523 | C | 523 | 21 | 1.0 | 168168 | 2 | AC012283 | AC012283 | Homo sapi | 596 | 21 | 1.0 | 194509 | 2 | AC107234 | AC107234 | Mus muscu |
| 524 | C | 524 | 21 | 1.0 | 168954 | 2 | AC099706 | AC099706 | Mus muscu | 597 | 21 | 1.0 | 194860 | 2 | AC113940 | AC113940 | Mus muscu |
| 525 | C | 525 | 21 | 1.0 | 169101 | 2 | AC008317 | AC008317 | Drosophill | 598 | 21 | 1.0 | 194994 | 9 | AC108690 | AC108690 | Homo sapi |
| 526 | C | 526 | 21 | 1.0 | 169311 | 3 | AC123805 | AC123805 | Mus muscu | 599 | 21 | 1.0 | 194985 | 10 | AL627103 | AL627103 | Mouse DNA |
| 527 | C | 527 | 21 | 1.0 | 169354 | 2 | AC099598 | AC099598 | Mus muscu | 600 | 21 | 1.0 | 196216 | 2 | AC019228 | AC019228 | Homo sapi |
| 528 | C | 528 | 21 | 1.0 | 170064 | 2 | AC095307 | AC095307 | Rattus no | 601 | 21 | 1.0 | 196235 | 9 | AC109810 | AC109810 | Homo sapi |
| 529 | C | 529 | 21 | 1.0 | 170347 | 3 | AC104054 | AC104054 | Drosophill | 602 | 21 | 1.0 | 196828 | 2 | AC127709 | AC127709 | Rattus no |
| 530 | C | 530 | 21 | 1.0 | 170443 | 9 | AC011086 | AC011086 | Rattus no | 603 | 21 | 1.0 | 197345 | 2 | AC110248 | AC110248 | Homo sapi |
| 531 | C | 531 | 21 | 1.0 | 170491 | 2 | AC019105 | AC019105 | Homo sapi | 604 | 21 | 1.0 | 197631 | 2 | AC098511 | AC098511 | Rattus no |
| 532 | C | 532 | 21 | 1.0 | 170711 | 2 | AC033020 | AC033020 | Homo sapi | 605 | 21 | 1.0 | 197797 | 2 | AC012514 | AC012514 | Homo sapi |
| 533 | C | 533 | 21 | 1.0 | 170736 | 2 | AC013658 | AC013658 | Homo sapi | 606 | 21 | 1.0 | 198427 | 8 | ATCHRIV52 | ATCHRIV52 | Homo sapi |
| 534 | C | 534 | 21 | 1.0 | 171272 | 2 | AC116684 | AC116684 | Mus muscu | 607 | 21 | 1.0 | 199031 | 2 | AC110416 | AC110416 | Rattus no |
| 535 | C | 535 | 21 | 1.0 | 171452 | 9 | AP000476 | AP000476 | Homo sapi | 608 | 21 | 1.0 | 199052 | 2 | AC122962 | AC122962 | Rattus no |
| 536 | C | 536 | 21 | 1.0 | 172191 | 2 | AL845505 | AL845505 | Mus muscu | 609 | 21 | 1.0 | 199103 | 2 | AC123813 | AC123813 | Mus muscu |
| 537 | C | 537 | 21 | 1.0 | 172649 | 2 | AC109744 | AC109744 | Rattus no | 610 | 21 | 1.0 | 199254 | 2 | AC021144 | AC021144 | Homo sapi |
| 538 | C | 538 | 21 | 1.0 | 173142 | 2 | AC091897 | AC091897 | Homo sapi | 611 | 21 | 1.0 | 199484 | 2 | AC129057 | AC129057 | Rattus no |
| 539 | C | 539 | 21 | 1.0 | 173295 | 9 | AC099744 | AC099744 | Papio cyn | 612 | 21 | 1.0 | 199579 | 2 | AC117899 | AC117899 | Rattus no |
| 540 | C | 540 | 21 | 1.0 | 173399 | 2 | AC112985 | AC112985 | Mus muscu | 613 | 21 | 1.0 | 199579 | 9 | AC093536 | AC093536 | Homo sapi |
| 541 | C | 541 | 21 | 1.0 | 173349 | 2 | AC105504 | AC105504 | Rattus no | 614 | 21 | 1.0 | 201724 | 2 | AC119506 | AC119506 | Rattus no |
| 542 | C | 542 | 21 | 1.0 | 174678 | 9 | AL450332 | AL450332 | Human DNA | 615 | 21 | 1.0 | 201879 | 2 | AC122011 | AC122011 | Mus muscu |
| 543 | C | 543 | 21 | 1.0 | 176036 | 3 | AC009258 | AC009258 | Drosophill | 616 | 21 | 1.0 | 203713 | 2 | AC084725 | AC084725 | Mus muscu |
| 544 | C | 544 | 21 | 1.0 | 177104 | 10 | AC091237 | AC091237 | Mus muscu | 617 | 21 | 1.0 | 204192 | 2 | AC121396 | AC121396 | Rattus no |
| 545 | C | 545 | 21 | 1.0 | 177404 | 2 | AC108296 | AC108296 | Rattus no | 618 | 21 | 1.0 | 204340 | 2 | AC019238 | AC019238 | Homo sapi |
| 546 | C | 546 | 21 | 1.0 | 178245 | 2 | AC107343 | AC107343 | Rattus no | 619 | 21 | 1.0 | 204959 | 2 | AC103945 | AC103945 | Mus muscu |
| 547 | C | 547 | 21 | 1.0 | 178273 | 2 | AC005308 | AC005308 | plasmodiu | 620 | 21 | 1.0 | 208405 | 2 | AC079477 | AC079477 | Mus muscu |
| 548 | C | 548 | 21 | 1.0 | 178452 | 2 | AC024672 | AC024672 | Homo sapi | 621 | 21 | 1.0 | 209815 | 2 | ALB45290 | ALB45290 | Mus muscu |
| 549 | C | 549 | 21 | 1.0 | 178561 | 9 | AC097685 | AC097685 | Rattus no | 622 | 21 | 1.0 | 211542 | 2 | AL672182 | AL672182 | Mus muscu |
| 550 | C | 550 | 21 | 1.0 | 178756 | 2 | AC104420 | AC104420 | Homo sapi | 623 | 21 | 1.0 | 211571 | 2 | AC125765 | AC125765 | Rattus no |
| 551 | C | 551 | 21 | 1.0 | 178942 | 9 | AL357520 | AL357520 | Human DNA | 624 | 21 | 1.0 | 211871 | 2 | AL845293 | AL845293 | Mus muscu |
| 552 | C | 552 | 21 | 1.0 | 180340 | 2 | AC016233 | AC016233 | Homo sapi | 625 | 21 | 1.0 | 213216 | 9 | HS399M15 | HS399M15 | Homo sapi |
| 553 | C | 553 | 21 | 1.0 | 180340 | 2 | AC016233 | AC016233 | Homo sapi | 626 | 21 | 1.0 | 214267 | 2 | AL844881 | AL844881 | Homo sapi |
| 554 | C | 554 | 21 | 1.0 | 180572 | 2 | AC013894 | AC013894 | Rattus no | 627 | 21 | 1.0 | 214727 | 2 | AC094443 | AC094443 | Mus muscu |
| 555 | C | 555 | 21 | 1.0 | 181365 | 2 | AC131194 | AC131194 | Rattus no | 628 | 21 | 1.0 | 215267 | 2 | AC079432 | AC079432 | Mus muscu |
| 556 | C | 556 | 21 | 1.0 | 181712 | 2 | AP001103 | AP001103 | Homo sapi | 629 | 21 | 1.0 | 215267 | 2 | AC021477 | AC021477 | Mus muscu |
| 557 | C | 557 | 21 | 1.0 | 182052 | 2 | AC119476 | AC119476 | Rattus no | 630 | 21 | 1.0 | 216120 | 2 | AL833780 | AL833780 | Mus muscu |
| 558 | C | 558 | 21 | 1.0 | 182314 | 2 | AC013670 | AC013670 | Homo sapi | 631 | 21 | 1.0 | 216671 | 2 | AC126597 | AC126597 | Rattus no |
| 559 | C | 559 | 21 | 1.0 | 183032 | 2 | AC109103 | AC109103 | Rattus no | 632 | 21 | 1.0 | 217634 | 2 | AC121979 | AC121979 | Mus muscu |
| 560 | C | 560 | 21 | 1.0 | 183099 | 2 | AC025288 | AC025288 | Homo sapi | 633 | 21 | 1.0 | 219491 | 2 | AC092256 | AC092256 | Mus muscu |
| 561 | C | 561 | 21 | 1.0 | 183658 | 2 | AC033021 | AC033021 | Homo sapi | 634 | 21 | 1.0 | 220159 | 2 | AC124507 | AC124507 | Mus muscu |
| 562 | C | 562 | 21 | 1.0 | 183690 | 9 | AC091180 | AC091180 | Homo sapi | 635 | 21 | 1.0 | 221669 | 2 | AC113026 | AC113026 | Mus muscu |
| 563 | C | 563 | 21 | 1.0 | 184013 | 2 | AC112632 | AC112632 | Rattus no | 636 | 21 | 1.0 | 222281 | 2 | AC096353 | AC096353 | Rattus no |
| 564 | C | 564 | 21 | 1.0 | 184284 | 2 | AC025582 | AC025582 | Mus muscu | 637 | 21 | 1.0 | 222826 | 2 | AC124728 | AC124728 | Mus muscu |
| 565 | C | 565 | 21 | 1.0 | 184531 | 2 | AC112677 | AC112677 | Mus muscu | 638 | 21 | 1.0 | 224788 | 9 | AC011501 | AC011501 | Homo sapi |
| 566 | C | 566 | 21 | 1.0 | 185108 | 2 | AC118835 | AC118835 | Rattus no | 639 | 21 | 1.0 | 226803 | 2 | AC123057 | AC123057 | Mus muscu |
| 567 | C | 567 | 21 | 1.0 | 187064 | 9 | AC011476 | AC011476 | Homo sapi | 640 | 21 | 1.0 | 227813 | 2 | AL772225 | AL772225 | Mus muscu |
| 568 | C | 568 | 21 | 1.0 | 187505 | 2 | AC121401 | AC121401 | Rattus no | 641 | 21 | 1.0 | 229037 | 2 | AC097323 | AC097323 | Mus muscu |
| 569 | C | 569 | 21 | 1.0 | 187561 | 2 | AC130850 | AC130850 | Rattus no | 642 | 21 | 1.0 | 232278 | 2 | AC087876 | AC087876 | Mus muscu |
| 570 | C | 570 | 21 | 1.0 | 187921 | 30 | AC096395 | AC096395 | Homo sapi | 643 | 21 | 1.0 | 233129 | 2 | AC094132 | AC094132 | Mus muscu |
| 571 | C | 571 | 21 | 1.0 | 188129 | 2 | AC095782 | AC095782 | Rattus no | 644 | 21 | 1.0 | 233775 | 2 | AC129336 | AC129336 | Mus muscu |
| 572 | C | 572 | 21 | 1.0 | 188143 | 2 | AL671872 | AL671872 | Mus muscu | 645 | 21 | 1.0 | 234960 | 2 | AC115675 | AC115675 | Rattus no |
| 573 | C | 573 | 21 | 1.0 | 188260 | 2 | AC110730 | AC110730 | Mus muscu | 646 | 21 | 1.0 | 235571 | 2 | AC128364 | AC128364 | Rattus no |
| 574 | C | 574 | 21 | 1.0 | 188644 | 9 | AC090518 | AC090518 | Homo sapi | 647 | 21 | 1.0 | 235948 | 2 | AC099574 | AC099574 | Mus muscu |
| 575 | C | 575 | 21 | 1.0 | 188818 | 2 | AC114019 | AC114019 | Rattus no | 648 | 21 | 1.0 | 236697 | 2 | AC096364 | AC096364 | Rattus no |
| 576 | C | 576 | 21 | 1.0 | 189940 | 2 | AC051647 | AC051647 | Homo sapi | 649 | 21 | 1.0 | 238528 | 2 | AC098896 | AC098896 | Rattus no |

| | | | | | | | | | | | | | |
|-------|----|-----|--------|----|------------|---------------------|-----|----|-----|------|----|------------|---------------------|
| c 650 | 21 | 1.0 | 238593 | 2 | AC093343 | Mus muscu | 723 | 20 | 1.0 | 1354 | 9 | BC001284 | BC001284 Homo sapi |
| c 651 | 21 | 1.0 | 239927 | 2 | AL731735 | Mus muscu | 724 | 20 | 1.0 | 1360 | 8 | AT292460 | AT292460 Arabidops |
| c 652 | 21 | 1.0 | 240809 | 2 | AC096352 | AC096352 | 725 | 20 | 1.0 | 1366 | 8 | AF339740 | AF339740 Vitis vin |
| c 653 | 21 | 1.0 | 242677 | 2 | AC113257 | Rattus no | 726 | 20 | 1.0 | 1436 | 8 | BC020889 | BC020889 Homo sapi |
| c 654 | 21 | 1.0 | 242677 | 2 | AC113257 | Rattus no | 727 | 20 | 1.0 | 1446 | 8 | AY093053 | AY093053 Arabidops |
| c 655 | 21 | 1.0 | 245383 | 2 | AL161653 | Homo sapi | 728 | 20 | 1.0 | 1451 | 8 | BC009205 | BC009205 Homo sapi |
| c 656 | 21 | 1.0 | 250925 | 2 | AC123053 | Mus muscu | 729 | 20 | 1.0 | 1463 | 3 | AY051872 | AY051872 Drosophill |
| c 657 | 21 | 1.0 | 251202 | 2 | AC091902 | AC091902 Homo sapi | 730 | 20 | 1.0 | 1469 | 8 | BC028026 | BC028026 Homo sapi |
| c 658 | 21 | 1.0 | 253305 | 3 | PFMAL137 | AL034559 Plasmodiu | 731 | 20 | 1.0 | 1504 | 8 | AY087356 | AY087356 Arabidops |
| c 659 | 21 | 1.0 | 254993 | 2 | AC078885 | Mus muscu | 732 | 20 | 1.0 | 1526 | 9 | AF131852 | AF131852 Homo sapi |
| c 660 | 21 | 1.0 | 256172 | 2 | AC005139 | AC005139 Plasmodiu | 733 | 20 | 1.0 | 1548 | 9 | AF068262 | AF068262 Drosophill |
| c 661 | 21 | 1.0 | 258024 | 2 | AC084046 | AC084046 Trypanoso | 734 | 20 | 1.0 | 1587 | 3 | AF067206 | AF067206 Drosophill |
| c 662 | 21 | 1.0 | 258456 | 2 | AC118802 | Rattus no | 735 | 20 | 1.0 | 1590 | 3 | AF065473 | AF065473 Drosophill |
| c 663 | 21 | 1.0 | 289090 | 2 | AE003424 | AE003424 Drosophill | 736 | 20 | 1.0 | 1634 | 8 | AY007108 | AY007108 Homo sapi |
| c 664 | 21 | 1.0 | 293431 | 2 | PFMAL13B4 | AL049181 Plasmodiu | 737 | 20 | 1.0 | 1649 | 8 | HBR132581 | AJ132581 Hevea bra |
| c 665 | 21 | 1.0 | 310001 | 9 | AF178030 | AF178030 Homo sapi | 738 | 20 | 1.0 | 1651 | 8 | HBR132580 | HBR132580 Hevea bra |
| c 666 | 21 | 1.0 | 310779 | 2 | AC005140 | AC005140 Plasmodiu | 739 | 20 | 1.0 | 1653 | 8 | B3SAMC1G | BC034430 Homo sapi |
| c 667 | 21 | 1.0 | 318221 | 2 | PFMAL13B3 | AL049184 Plasmodiu | 740 | 20 | 1.0 | 1725 | 9 | BC034430 | BC034430 Homo sapi |
| c 668 | 21 | 1.0 | 329362 | 3 | AE003681 | AE003681 Drosophill | 741 | 20 | 1.0 | 1734 | 9 | BC016868 | BC016868 Homo sapi |
| c 669 | 21 | 1.0 | 340000 | 9 | AP001690 | AP001690 Homo sapi | 742 | 20 | 1.0 | 1737 | 9 | HSB800182 | AL049397 Homo sapi |
| c 670 | 20 | 1.0 | 340000 | 9 | AP001690 | M29110 D.discoiden | 743 | 20 | 1.0 | 1750 | 10 | BC018512 | BC018512 Mus muscu |
| c 671 | 20 | 1.0 | 250 | 11 | G31057 | G31057 SHGC-19072 | 744 | 20 | 1.0 | 1833 | 10 | BC004687 | BC004687 Mus muscu |
| c 672 | 20 | 1.0 | 250 | 11 | G33473 | AR099706 Sequence | 745 | 20 | 1.0 | 1855 | 9 | AB060855 | AB060855 Macaca fa |
| c 673 | 20 | 1.0 | 295 | 6 | AR099706 | 122435 Sequence | 746 | 20 | 1.0 | 1869 | 8 | ATSKIN2 | AF053369 Homo sapi |
| c 674 | 20 | 1.0 | 295 | 6 | 122435 | 122435 Sequence | 747 | 20 | 1.0 | 1913 | 9 | AF053369 | AF053369 Homo sapi |
| c 675 | 20 | 1.0 | 295 | 11 | G62176 | AR091702 Sequence | 748 | 20 | 1.0 | 1956 | 8 | AF036328 | AF036328 Arabidops |
| c 676 | 20 | 1.0 | 362 | 6 | AR099702 | 122431 Sequence | 749 | 20 | 1.0 | 1981 | 8 | HSB802144 | AL137654 Homo sapi |
| c 677 | 20 | 1.0 | 362 | 6 | 122431 | Y15594 Drosophilla | 750 | 20 | 1.0 | 2026 | 3 | AF071409 | AF071409 Plasmodiu |
| c 678 | 20 | 1.0 | 366 | 3 | DMACR2 | AR177413 Sequence | 751 | 20 | 1.0 | 2051 | 9 | AK000527 | AK000527 Homo sapi |
| c 679 | 20 | 1.0 | 387 | 6 | AR177413 | AE511233 Sequence | 752 | 20 | 1.0 | 2128 | 3 | AF027825 | AF027825 Plasmodiu |
| c 680 | 20 | 1.0 | 402 | 4 | AE511233 | AX381356 Sequence | 753 | 20 | 1.0 | 2134 | 9 | AK025213 | AK025213 Homo sapi |
| c 681 | 20 | 1.0 | 410 | 6 | AX381356 | AE427567 B1K1nla e | 754 | 20 | 1.0 | 2198 | 9 | AK025213 | AY093170 Homo sapi |
| c 682 | 20 | 1.0 | 434 | 8 | AE427567 | AE427567 Tetrahel | 755 | 20 | 1.0 | 2209 | 8 | BC025940 | BC025940 Mus muscu |
| c 683 | 20 | 1.0 | 443 | 8 | AE427567 | AE427567 Tetrahel | 756 | 20 | 1.0 | 2228 | 10 | BC006871 | BC006871 Mus muscu |
| c 684 | 20 | 1.0 | 444 | 8 | AE427567 | AE427567 Tetrahel | 757 | 20 | 1.0 | 2256 | 10 | BC006871 | BC006871 Mus muscu |
| c 685 | 20 | 1.0 | 447 | 8 | AE427570 | AF427570 B1K1nla 1 | 758 | 20 | 1.0 | 2271 | 5 | AF014367 | AF014367 Homo sapi |
| c 686 | 20 | 1.0 | 447 | 8 | AE427570 | AF427570 Tetrahel | 759 | 20 | 1.0 | 2320 | 5 | HSB420511 | HSB420511 Homo sapi |
| c 687 | 20 | 1.0 | 447 | 8 | AE427570 | AF427570 Tetrahel | 760 | 20 | 1.0 | 2429 | 9 | AF139207 | AF139207 Arabidops |
| c 688 | 20 | 1.0 | 450 | 8 | AE427571 | AE427571 B1K1nla 1 | 761 | 20 | 1.0 | 2407 | 9 | BC018361 | BC018361 Homo sapi |
| c 689 | 20 | 1.0 | 452 | 8 | AE427543 | AE427543 Aphnocal | 762 | 20 | 1.0 | 2564 | 10 | BC011270 | BC011270 Mus muscu |
| c 690 | 20 | 1.0 | 453 | 8 | AE427574 | AE427574 Tetrahel | 763 | 20 | 1.0 | 2586 | 8 | AY034972 | AY034972 Arabidops |
| c 691 | 20 | 1.0 | 459 | 8 | AE427542 | AE427542 Aphnocal | 764 | 20 | 1.0 | 2606 | 3 | DDICAR | L09637 Dictyostell |
| c 692 | 20 | 1.0 | 474 | 6 | AX284811 | AX284811 Sequence | 765 | 20 | 1.0 | 2687 | 8 | BPE289770 | BC024282 Homo sapi |
| c 693 | 20 | 1.0 | 517 | 6 | AX400630 | AX400630 Sequence | 766 | 20 | 1.0 | 2787 | 9 | BC011700 | BC011700 Homo sapi |
| c 694 | 20 | 1.0 | 537 | 10 | BC021776 | BC021776 Mus muscu | 767 | 20 | 1.0 | 2833 | 3 | PFATUBB | PFATUBB Homo sapi |
| c 695 | 20 | 1.0 | 554 | 6 | AX381524 | AX381524 Sequence | 768 | 20 | 1.0 | 2845 | 9 | BC001427 | BC001427 Homo sapi |
| c 696 | 20 | 1.0 | 686 | 8 | CA025718 | U25718 Candida alb | 769 | 20 | 1.0 | 2851 | 8 | AY065425 | AY065425 Arabidops |
| c 697 | 20 | 1.0 | 722 | 11 | PM2A7G | AL684684 Penicillii | 770 | 20 | 1.0 | 2887 | 10 | BC021881 | BC021881 Mus muscu |
| c 698 | 20 | 1.0 | 733 | 9 | HSB332860 | AJ332860 Homo sapi | 771 | 20 | 1.0 | 2894 | 6 | AX003083 | AX003083 Sequence |
| c 699 | 20 | 1.0 | 800 | 8 | SCSCD26A | X17118 Yeast SCD26 | 772 | 20 | 1.0 | 2894 | 6 | E35444 | E35444 SBEEN1 pol |
| c 700 | 20 | 1.0 | 839 | 10 | AF012282 | X17118 Yeast SCD26 | 773 | 20 | 1.0 | 2908 | 9 | RATPMP70 | BC024282 Homo sapi |
| c 701 | 20 | 1.0 | 853 | 9 | BC022815 | BC022815 Homo sapi | 774 | 20 | 1.0 | 2923 | 8 | AP213696 | AP213696 Nicotiana |
| c 702 | 20 | 1.0 | 879 | 9 | BC004466 | BC004466 Homo sapi | 775 | 20 | 1.0 | 2959 | 6 | AX449346 | AX449346 Sequence |
| c 703 | 20 | 1.0 | 898 | 3 | AY113599 | AY113599 Drosophill | 776 | 20 | 1.0 | 3263 | 5 | GDVHC | X95552 G.domesticu |
| c 704 | 20 | 1.0 | 996 | 9 | BC001864 | BC001864 Homo sapi | 777 | 20 | 1.0 | 3324 | 10 | RATPMP70 | BC024282 Homo sapi |
| c 705 | 20 | 1.0 | 998 | 8 | AY050914 | AY050914 Arabidops | 778 | 20 | 1.0 | 3338 | 8 | AY064013 | AY064013 Arabidops |
| c 706 | 20 | 1.0 | 1033 | 3 | PFAR459C | M83792 Plasmodiu | 779 | 20 | 1.0 | 3340 | 10 | AF159455 | AF159455 Mus muscu |
| c 707 | 20 | 1.0 | 1061 | 3 | AF001981 | AF001981 Dictyoste | 780 | 20 | 1.0 | 3416 | 8 | ATMARKN10A | M93023 Arabidops |
| c 708 | 20 | 1.0 | 1064 | 10 | HSAN06941 | AJ06941 Homo sapi | 781 | 20 | 1.0 | 3454 | 8 | HSB803569 | HSB803569 Homo sapi |
| c 709 | 20 | 1.0 | 1080 | 10 | MPPS00RFS | X85177 M.musculus | 782 | 20 | 1.0 | 3763 | 6 | E17067 | E17067 Human mRNa |
| c 710 | 20 | 1.0 | 1094 | 4 | AB049188 | AB049188 Equus cab | 783 | 20 | 1.0 | 3763 | 6 | E17068 | E17068 Human cDNA |
| c 711 | 20 | 1.0 | 1104 | 6 | AX449318 | AX449318 Sequence | 784 | 20 | 1.0 | 3763 | 6 | AB003062 | AB003062 Homo sapi |
| c 712 | 20 | 1.0 | 1104 | 6 | AX411609 | AX411609 Sequence | 785 | 20 | 1.0 | 3796 | 10 | RNFIBN1 | X05831 Rat fibron |
| c 713 | 20 | 1.0 | 1109 | 8 | AF350937 | AF350937 Prunus ce | 786 | 20 | 1.0 | 3882 | 9 | BC028385 | BC028385 Homo sapi |
| c 714 | 20 | 1.0 | 1227 | 8 | AY070073 | AY070073 Arabidops | 787 | 20 | 1.0 | 3971 | 3 | AF310888 | AF310888 Dictyoste |
| c 715 | 20 | 1.0 | 1231 | 5 | BC015999 | BC015999 Homo sapi | 788 | 20 | 1.0 | 4382 | 3 | PFSC04084 | PFSC04084 Plasmodiu |
| c 716 | 20 | 1.0 | 1284 | 5 | AF180887 | AF180887 Dnlo rer | 789 | 20 | 1.0 | 4560 | 3 | HSB803330 | HSB803330 Homo sapi |
| c 717 | 20 | 1.0 | 1291 | 6 | A26098 | A26098 P.falciparu | 790 | 20 | 1.0 | 4845 | 2 | AC015305 | AC015305 Drosophill |
| c 718 | 20 | 1.0 | 1291 | 6 | A32873 | AL833755 Homo sapi | 791 | 20 | 1.0 | 5036 | 2 | HSB802517 | AL161952 Homo sapi |
| c 719 | 20 | 1.0 | 1332 | 9 | HSB805068 | AL833755 Homo sapi | 792 | 20 | 1.0 | 5310 | 6 | HSB802884 | AL512695 Homo sapi |
| c 720 | 20 | 1.0 | 1336 | 9 | PFAR493427 | AJ493427 Plasmodiu | 793 | 20 | 1.0 | 5814 | 6 | AX431123 | AX431123 Sequence |
| c 721 | 20 | 1.0 | 1350 | 9 | BC015907 | BC015907 Homo sapi | 794 | 20 | 1.0 | 5930 | 5 | HSB803271 | AB831965 Homo sapi |
| c 722 | 20 | 1.0 | 1351 | 3 | AY071383 | AY071383 Drosophill | 795 | 20 | 1.0 | 5995 | 5 | AB032197 | AB032197 Gallus ga |

| | | | | | | | | | | | | | | | |
|-------|----|-----|-------|----|------------|------------|------------|-------|----|-----|--------|----|-----------|-----------|--------------------|
| C 796 | 20 | 1.0 | 6123 | 9 | HSMB04590 | AL833277 | Homo sapi | C 869 | 20 | 1.0 | 74313 | 8 | AP004977 | AP004977 | Lotus jap |
| C 797 | 20 | 1.0 | 7138 | 6 | AX323841 | AX323841 | Sequence | C 870 | 20 | 1.0 | 75756 | 8 | AP0068601 | AP0068601 | Homo sapi |
| C 798 | 20 | 1.0 | 11102 | 6 | AF324888 | AF324888 | Homo sapi | C 871 | 20 | 1.0 | 77636 | 8 | AB077832 | AB077832 | Arabidops |
| C 799 | 20 | 1.0 | 12029 | 3 | AE001381 | AE001381 | Plasmodi | C 872 | 20 | 1.0 | 78578 | 2 | PFMAL159 | PFMAL159 | Plasmodi |
| C 800 | 20 | 1.0 | 12029 | 3 | AE001412 | AE001412 | Plasmodi | C 873 | 20 | 1.0 | 78770 | 2 | AC034230 | AC034230 | Homo sapi |
| C 801 | 20 | 1.0 | 12029 | 3 | AE001427 | AE001427 | Plasmodi | C 874 | 20 | 1.0 | 79457 | 2 | AC098396 | AC098396 | Rattus no |
| C 802 | 20 | 1.0 | 12229 | 3 | AE001377 | AE001377 | Plasmodi | C 875 | 20 | 1.0 | 79769 | 2 | AC115678 | AC115678 | Dictyoste |
| C 803 | 20 | 1.0 | 12355 | 2 | AC116987 | AC116987 | Dictyoste | C 876 | 20 | 1.0 | 80030 | 2 | AC097127 | AC097127 | Rattus no |
| C 804 | 20 | 1.0 | 14202 | 3 | AF324889S3 | AF324889S3 | Homo sapi | C 877 | 20 | 1.0 | 80845 | 2 | AC098831 | AC098831 | Homo sapi |
| C 805 | 20 | 1.0 | 14613 | 3 | AE001396 | AE001396 | Plasmodi | C 878 | 20 | 1.0 | 81117 | 9 | AC016898 | AC016898 | Homo sapi |
| C 806 | 20 | 1.0 | 15342 | 3 | AE001420 | AE001420 | Plasmodi | C 879 | 20 | 1.0 | 82411 | 8 | AC068809 | AC068809 | Genomic s |
| C 807 | 20 | 1.0 | 15387 | 1 | SC5C11 | SC5C11 | Streptomy | C 880 | 20 | 1.0 | 82875 | 8 | AC063482 | AC063482 | Genomic s |
| C 808 | 20 | 1.0 | 15903 | 2 | AC115612 | AC115612 | Dictyoste | C 881 | 20 | 1.0 | 84162 | 8 | T22J18 | T22J18 | Arabidops |
| C 809 | 20 | 1.0 | 15903 | 2 | AC115612 | AC115612 | Dictyoste | C 882 | 20 | 1.0 | 84763 | 9 | AL356134 | AL356134 | Human DNA |
| C 810 | 20 | 1.0 | 19369 | 2 | AC131308 | AC131308 | Homo sapi | C 883 | 20 | 1.0 | 89703 | 2 | AC107601 | AC107601 | Rattus no |
| C 811 | 20 | 1.0 | 21882 | 10 | AL606782 | AL606782 | Mouse DNA | C 884 | 20 | 1.0 | 90289 | 2 | AC109943 | AC109943 | Rattus no |
| C 812 | 20 | 1.0 | 23324 | 2 | AC020471 | AC020471 | Drosophi | C 885 | 20 | 1.0 | 91013 | 2 | PFMAL1_3 | PFMAL1_3 | Continuation (4 of |
| C 813 | 20 | 1.0 | 24053 | 9 | ATY12227 | ATY12227 | Arabidops | C 886 | 20 | 1.0 | 92633 | 2 | AC005917 | AC005917 | Arabidops |
| C 814 | 20 | 1.0 | 24800 | 9 | AF448221 | AF448221 | Homo sapi | C 887 | 20 | 1.0 | 93282 | 8 | AC097353 | AC097353 | Rattus no |
| C 815 | 20 | 1.0 | 25117 | 2 | AC116102 | AC116102 | Dictyoste | C 888 | 20 | 1.0 | 93329 | 2 | AC116967 | AC116967 | Dictyoste |
| C 816 | 20 | 1.0 | 25339 | 3 | AB017765 | AB017765 | Trypanoso | C 889 | 20 | 1.0 | 93431 | 2 | AC107463 | AC107463 | Homo sapi |
| C 817 | 20 | 1.0 | 25499 | 2 | AC017653 | AC017653 | Drosophi | C 890 | 20 | 1.0 | 93681 | 9 | HS424E5 | HS424E5 | Human DNA |
| C 818 | 20 | 1.0 | 27744 | 2 | AC019881 | AC019881 | Drosophi | C 891 | 20 | 1.0 | 93805 | 2 | AC108107 | AC108107 | Homo sapi |
| C 819 | 20 | 1.0 | 27934 | 9 | AC106717 | AC106717 | Homo sapi | C 892 | 20 | 1.0 | 94232 | 9 | AC117861 | AC117861 | Rattus no |
| C 820 | 20 | 1.0 | 29001 | 2 | AC018796 | AC018796 | Drosophi | C 893 | 20 | 1.0 | 95175 | 2 | ATF18P9 | ATF18P9 | Rattus no |
| C 821 | 20 | 1.0 | 30220 | 2 | AC019816 | AC019816 | Drosophi | C 894 | 20 | 1.0 | 95996 | 8 | AC103251 | AC103251 | Rattus no |
| C 822 | 20 | 1.0 | 31520 | 9 | HSX1ST1 | HSX1ST1 | Human X413 | C 895 | 20 | 1.0 | 97044 | 8 | AC005698 | AC005698 | Genomic s |
| C 823 | 20 | 1.0 | 31812 | 8 | DMBNA3K23 | DMBNA3K23 | S.pombe | C 896 | 20 | 1.0 | 99680 | 2 | AC099769 | AC099769 | Homo sapi |
| C 824 | 20 | 1.0 | 32391 | 3 | AC003690 | AC003690 | Human Chr | C 897 | 20 | 1.0 | 99964 | 2 | RN364106 | RN364106 | Homo sapi |
| C 825 | 20 | 1.0 | 34052 | 9 | AC0129092 | AC0129092 | Medicago | C 898 | 20 | 1.0 | 100000 | 9 | AB020859 | AB020859 | Rattus no |
| C 826 | 20 | 1.0 | 34096 | 2 | AC116958 | AC116958 | Dictyoste | C 899 | 20 | 1.0 | 100000 | 9 | AB020854 | AB020854 | Homo sapi |
| C 827 | 20 | 1.0 | 35119 | 9 | HS27C3 | HS27C3 | Human DNA | C 900 | 20 | 1.0 | 100000 | 9 | AB020872 | AB020872 | Homo sapi |
| C 828 | 20 | 1.0 | 35700 | 9 | AC013004 | AC013004 | Drosophi | C 901 | 20 | 1.0 | 100000 | 9 | AP000210 | AP000210 | Homo sapi |
| C 829 | 20 | 1.0 | 36576 | 2 | AC011529 | AC011529 | Homo sapi | C 902 | 20 | 1.0 | 100000 | 17 | AP000132 | AP000132 | Homo sapi |
| C 830 | 20 | 1.0 | 37604 | 9 | AC011548 | AC011548 | Homo sapi | C 903 | 20 | 1.0 | 100282 | 8 | AC096298 | AC096298 | Rattus no |
| C 831 | 20 | 1.0 | 38351 | 6 | AX244215 | AX244215 | Sequence | C 904 | 20 | 1.0 | 101213 | 3 | AC113851 | AC113851 | Rattus no |
| C 832 | 20 | 1.0 | 38653 | 9 | AC012633 | AC012633 | Homo sapi | C 905 | 20 | 1.0 | 101644 | 8 | AC006827 | AC006827 | Arabidops |
| C 833 | 20 | 1.0 | 40965 | 9 | AC090631 | AC090631 | Homo sapi | C 906 | 20 | 1.0 | 102326 | 2 | AC120734 | AC120734 | Rattus no |
| C 834 | 20 | 1.0 | 41135 | 9 | AF217651 | AF217651 | Drosophi | C 907 | 20 | 1.0 | 102422 | 2 | AC129557 | AC129557 | Mus muscu |
| C 835 | 20 | 1.0 | 45142 | 3 | RN363K5 | RN363K5 | Rattus no | C 908 | 20 | 1.0 | 103922 | 2 | AC096546 | AC096546 | Homo sapi |
| C 836 | 20 | 1.0 | 47791 | 2 | AC115593 | AC115593 | Dictyoste | C 909 | 20 | 1.0 | 103942 | 2 | AC091859 | AC091859 | Homo sapi |
| C 837 | 20 | 1.0 | 47791 | 2 | AC116305 | AC116305 | Dictyoste | C 910 | 20 | 1.0 | 103989 | 9 | HSDA109P7 | HSDA109P7 | Human DNA |
| C 838 | 20 | 1.0 | 47791 | 2 | AC116305 | AC116305 | Dictyoste | C 911 | 20 | 1.0 | 104320 | 2 | AC126914 | AC126914 | Rattus no |
| C 839 | 20 | 1.0 | 48352 | 9 | AC0101079 | AC0101079 | Mus muscu | C 912 | 20 | 1.0 | 104351 | 9 | AC0922419 | AC0922419 | Homo sapi |
| C 840 | 20 | 1.0 | 48871 | 8 | AC082643 | AC082643 | Arabidops | C 913 | 20 | 1.0 | 104786 | 2 | AC111859 | AC111859 | Rattus no |
| C 841 | 20 | 1.0 | 48891 | 8 | AC103694 | AC103694 | Homo sapi | C 914 | 20 | 1.0 | 104982 | 2 | AC005504 | AC005504 | Plasmodi |
| C 842 | 20 | 1.0 | 50711 | 2 | AC116551 | AC116551 | Dictyoste | C 915 | 20 | 1.0 | 105036 | 9 | AC073181 | AC073181 | Homo sapi |
| C 843 | 20 | 1.0 | 52514 | 2 | AC122764 | AC122764 | Mus muscu | C 916 | 20 | 1.0 | 105306 | 8 | ATP9D24 | ATP9D24 | Arabidops |
| C 844 | 20 | 1.0 | 54472 | 2 | AC116963 | AC116963 | Dictyoste | C 917 | 20 | 1.0 | 105797 | 9 | AC011345 | AC011345 | Homo sapi |
| C 845 | 20 | 1.0 | 56152 | 2 | AC100895 | AC100895 | Mus muscu | C 918 | 20 | 1.0 | 106039 | 9 | AL589786 | AL589786 | Human DNA |
| C 846 | 20 | 1.0 | 57455 | 2 | AC100895 | AC100895 | Mus muscu | C 919 | 20 | 1.0 | 106144 | 2 | AC116425 | AC116425 | Dictyoste |
| C 847 | 20 | 1.0 | 59004 | 9 | AC078882 | AC078882 | Dictyoste | C 920 | 20 | 1.0 | 106144 | 2 | AC116425 | AC116425 | Dictyoste |
| C 848 | 20 | 1.0 | 59427 | 9 | AC078882 | AC078882 | Dictyoste | C 921 | 20 | 1.0 | 106146 | 2 | AC116958 | AC116958 | Rattus no |
| C 849 | 20 | 1.0 | 59641 | 10 | D78344 | D78344 | Mouse DNA | C 922 | 20 | 1.0 | 106434 | 3 | AC117080 | AC117080 | Rattus no |
| C 850 | 20 | 1.0 | 61643 | 2 | AC021776 | AC021776 | Homo sapi | C 923 | 20 | 1.0 | 106638 | 2 | AC117983 | AC117983 | Rattus no |
| C 851 | 20 | 1.0 | 62001 | 2 | AC015206 | AC015206 | Drosophi | C 924 | 20 | 1.0 | 106983 | 2 | AC127409 | AC127409 | Rattus no |
| C 852 | 20 | 1.0 | 62497 | 2 | AC103865 | AC103865 | Homo sapi | C 925 | 20 | 1.0 | 107658 | 2 | AC119592 | AC119592 | Rattus no |
| C 853 | 20 | 1.0 | 63490 | 2 | AC104349 | AC104349 | Homo sapi | C 926 | 20 | 1.0 | 107658 | 2 | AC104067 | AC104067 | Homo sapi |
| C 854 | 20 | 1.0 | 64480 | 2 | AC118039 | AC118039 | Mus muscu | C 927 | 20 | 1.0 | 108346 | 9 | AC084298 | AC084298 | Homo sapi |
| C 855 | 20 | 1.0 | 64719 | 2 | AC128496 | AC128496 | Homo sapi | C 928 | 20 | 1.0 | 108729 | 2 | HS838L14 | HS838L14 | Homo sapi |
| C 856 | 20 | 1.0 | 65353 | 2 | AC100230 | AC100230 | Mus muscu | C 929 | 20 | 1.0 | 109280 | 2 | AC084053 | AC084053 | Mus muscu |
| C 857 | 20 | 1.0 | 65493 | 2 | AC099807 | AC099807 | Homo sapi | C 930 | 20 | 1.0 | 110000 | 2 | AC112264 | AC112264 | Continuation (3 of |
| C 858 | 20 | 1.0 | 65710 | 2 | AC102158 | AC102158 | Mus muscu | C 931 | 20 | 1.0 | 110000 | 2 | AC112266 | AC112266 | Mus muscu |
| C 859 | 20 | 1.0 | 65711 | 2 | AC025892 | AC025892 | Homo sapi | C 932 | 20 | 1.0 | 110000 | 2 | AC112266 | AC112266 | Mus muscu |
| C 860 | 20 | 1.0 | 66441 | 3 | PFMAL1P4 | PFMAL1P4 | Plasmodi | C 933 | 20 | 1.0 | 110000 | 2 | PFMAL1_1 | PFMAL1_1 | Continuation (2 of |
| C 861 | 20 | 1.0 | 67358 | 2 | AC108426 | AC108426 | Mus muscu | C 934 | 20 | 1.0 | 110000 | 2 | PFMAL1_1 | PFMAL1_1 | Continuation (2 of |
| C 862 | 20 | 1.0 | 67540 | 2 | AC079979 | AC079979 | Homo sapi | C 935 | 20 | 1.0 | 111298 | 2 | RN39A23 | RN39A23 | Plasmodi |
| C 863 | 20 | 1.0 | 68529 | 2 | AC101227 | AC101227 | Mus muscu | C 936 | 20 | 1.0 | 111298 | 2 | HS359615 | HS359615 | Human DNA |
| C 864 | 20 | 1.0 | 70201 | 2 | AC123638 | AC123638 | Mus muscu | C 937 | 20 | 1.0 | 112616 | 9 | AC025461 | AC025461 | Homo sapi |
| C 865 | 20 | 1.0 | 70667 | 2 | AC017306 | AC017306 | Drosophi | C 938 | 20 | 1.0 | 114630 | 2 | AC079628 | AC079628 | Homo sapi |
| C 866 | 20 | 1.0 | 70735 | 10 | AL663116 | AL663116 | Mouse DNA | C 939 | 20 | 1.0 | 114897 | 2 | AP003624 | AP003624 | Oryza sat |
| C 867 | 20 | 1.0 | 74331 | 2 | AP000575 | AP000575 | Homo sapi | C 940 | 20 | 1.0 | 115489 | 2 | AC117072 | AC117072 | Dictyoste |
| C 868 | 20 | 1.0 | 74089 | 9 | AL589762 | AL589762 | Human DNA | C 941 | 20 | 1.0 | 116636 | 3 | PFMAL3P3 | PFMAL3P3 | Plasmodi |

| | | | | | | |
|-------|----|------------|----|-----------|--------|------------|
| c 942 | 20 | 1.0 116596 | 3 | PFMAL3P3 | 298547 | Plasmodium |
| c 943 | 20 | 1.0 116747 | 2 | AC123329 | | |
| c 944 | 20 | 1.0 116897 | 2 | AC104644 | | |
| c 945 | 20 | 1.0 117074 | 2 | AC110940 | | |
| c 946 | 20 | 1.0 117585 | 2 | AC108608 | | |
| c 947 | 20 | 1.0 118588 | 9 | AL353774 | | |
| c 948 | 20 | 1.0 119452 | 2 | AC126302 | | |
| c 949 | 20 | 1.0 120492 | 3 | CEY43P8C | | |
| c 950 | 20 | 1.0 120592 | 2 | AL1714031 | | |
| c 951 | 20 | 1.0 120722 | 2 | AC111994 | | |
| c 952 | 20 | 1.0 120857 | 9 | AC116334 | | |
| c 953 | 20 | 1.0 121737 | 9 | AC008715 | | |
| c 954 | 20 | 1.0 122073 | 2 | AC010112 | | |
| c 955 | 20 | 1.0 122252 | 2 | AC109880 | | |
| c 956 | 20 | 1.0 122272 | 2 | AL359845 | | |
| c 957 | 20 | 1.0 122279 | 9 | AC119112 | | |
| c 958 | 20 | 1.0 122637 | 9 | AC006062 | | |
| c 959 | 20 | 1.0 123260 | 10 | AC087041 | | |
| c 960 | 20 | 1.0 123280 | 2 | AC117076 | | |
| c 961 | 20 | 1.0 124114 | 9 | HSJ858M2 | | |
| c 962 | 20 | 1.0 124230 | 8 | AC005259 | | |
| c 963 | 20 | 1.0 124321 | 10 | OSA245900 | | |
| c 964 | 20 | 1.0 125066 | 9 | AC074286 | | |
| c 965 | 20 | 1.0 125722 | 2 | AC015695 | | |
| c 966 | 20 | 1.0 127168 | 9 | HS116301 | | |
| c 967 | 20 | 1.0 127423 | 2 | AC117018 | | |
| c 968 | 20 | 1.0 130012 | 2 | AC120632 | | |
| c 969 | 20 | 1.0 130369 | 9 | AC074030 | | |
| c 970 | 20 | 1.0 130403 | 9 | AL158139 | | |
| c 971 | 20 | 1.0 131338 | 9 | AC099798 | | |
| c 972 | 20 | 1.0 131353 | 9 | HS508115 | | |
| c 973 | 20 | 1.0 131736 | 2 | AC021527 | | |
| c 974 | 20 | 1.0 131919 | 9 | AC008700 | | |
| c 975 | 20 | 1.0 132000 | 2 | AC116976 | | |
| c 976 | 20 | 1.0 132035 | 2 | AC125548 | | |
| c 977 | 20 | 1.0 132655 | 3 | AC010015 | | |
| c 978 | 20 | 1.0 132735 | 9 | AC107956 | | |
| c 979 | 20 | 1.0 133279 | 9 | AC012317 | | |
| c 980 | 20 | 1.0 133348 | 9 | AC114322 | | |
| c 981 | 20 | 1.0 133445 | 2 | AC073440 | | |
| c 982 | 20 | 1.0 134137 | 9 | HS3E5 | | |
| c 983 | 20 | 1.0 134319 | 2 | AC068673 | | |
| c 984 | 20 | 1.0 135123 | 9 | AC079950 | | |
| c 985 | 20 | 1.0 135439 | 2 | AC080065 | | |
| c 986 | 20 | 1.0 137060 | 2 | AC116867 | | |
| c 987 | 20 | 1.0 137077 | 9 | AC007562 | | |
| c 988 | 20 | 1.0 137942 | 10 | AL645745 | | |
| c 989 | 20 | 1.0 138151 | 2 | AC094070 | | |
| c 990 | 20 | 1.0 138264 | 9 | AC008722 | | |
| c 991 | 20 | 1.0 139736 | 2 | AC112779 | | |
| c 992 | 20 | 1.0 140133 | 2 | AC091050 | | |
| c 993 | 20 | 1.0 140583 | 9 | AC010978 | | |
| c 994 | 20 | 1.0 140702 | 2 | AC068420 | | |
| c 995 | 20 | 1.0 140952 | 2 | AC079982 | | |
| c 996 | 20 | 1.0 141790 | 9 | AL590636 | | |
| c 997 | 20 | 1.0 141899 | 9 | AC006334 | | |
| c 998 | 20 | 1.0 143331 | 9 | AC091214 | | |
| c 999 | 20 | 1.0 144515 | 9 | AC005831 | | |
| c1000 | 20 | 1.0 144882 | 2 | AC008726 | | |

ALIGNMENTS

| | | | | | | |
|------------|------------------------------------|-------------|-----|--------|-----|-------------|
| RESULT 1 | | | | | | |
| AR203335 | AR203335 | 2032 bp | DNA | linear | PAT | 20-JUN-2002 |
| LOCUS | Sequence 2 from patent US 6363365. | | | | | |
| DEFINITION | AR203335 | | | | | |
| ACCESSION | AR203335.1 | GI:21499698 | | | | |
| VERSION | | | | | | |
| KEYWORDS | | | | | | |
| SOURCE | Unknown. | | | | | |
| ORGANISM | Unknown. | | | | | |

| | | |
|-----------------------|---|---------------|
| REFERENCE | 1 (bases 1 to 2032) | Unclassified. |
| AUTHORS | Bisrup, A., Rosen, S.D., Tangemann, K. and Hemmerich, S. | |
| TITLE | Method of determining whether an agent modulates glycosyl | |
| JOURNAL | sulfotransferase-3 | |
| FEATURES | Patent: US 6363365-A 2 02-APR-2002; | |
| source | Location/Qualifiers | |
| BASE COUNT | 468 a 569 c 490 g 505 t | |
| ORIGIN | 1, 2032 | |
| Query Match | 100.0%; Score 2032; DB 6; Length 2032; | |
| Best local similarity | 100.0%; Pred. No. 0; | |
| Matches 2032; | Conservative 0; Mismatches 0; Indels 0; Gaps 0; | |
| 1 | GGCTCAGGCCAGAGTGCCTCCAGTCTGGGGGAAATGCTTCTATTGCTTCCAG | 60 |
| 1 | GGCTCAGGCCAGAGTGCCTCCAGTCTGGGGGAAATGCTTCTATTGCTTCCAG | 60 |
| 61 | CCACCTCAGAGAGTCCCAACCCCTGAGTCCAGAGTTAAAGTGTATTATTTCA | 120 |
| 61 | CCACCTCAGAGAGTCCCAACCCCTGAGTCCAGAGTTAAAGTGTATTATTTCA | 120 |
| 61 | CCACCTCAGAGAGTCCCAACCCCTGAGTCCAGAGTTAAAGTGTATTATTTCA | 120 |
| 121 | CAGCTTCTGAGGAGCGAGTCTTCTCAAGCCGCTCTGCAAGTCTTCCACTTCAG | 180 |
| 121 | CAGCTTCTGAGGAGCGAGTCTTCTCAAGCCGCTCTGCAAGTCTTCCACTTCAG | 180 |
| 121 | CAGCTTCTGAGGAGCGAGTCTTCTCAAGCCGCTCTGCAAGTCTTCCACTTCAG | 180 |
| 181 | AATGCTACGCTTAAATAATAGTCTGCTGTTCTGCTTCCAGATGGCATCTT | 240 |
| 181 | AATGCTACGCTTAAATAATAGTCTGCTGTTCTGCTTCCAGATGGCATCTT | 240 |
| 181 | AATGCTACGCTTAAATAATAGTCTGCTGTTCTGCTTCCAGATGGCATCTT | 240 |
| 241 | GGCTATTCCTTCCACATATGACGACACATACATGCTCTGCTATGAAAGGACAC | 300 |
| 241 | GGCTATTCCTTCCACATATGACGACACATACATGCTCTGCTATGAAAGGACAC | 300 |
| 241 | GGCTATTCCTTCCACATATGACGACACATACATGCTCTGCTATGAAAGGACAC | 300 |
| 301 | CGAGCGCATGACAGTCTGCTGCTCTTCTGCGGCTCTGCTCTTCTTGTGGGCA | 360 |
| 301 | CGAGCGCATGACAGTCTGCTGCTCTTCTGCGGCTCTGCTCTTCTTGTGGGCA | 360 |
| 301 | CGAGCGCATGACAGTCTGCTGCTCTTCTGCGGCTCTGCTCTTCTTGTGGGCA | 360 |
| 361 | GGCTTTTGGGAGACACCGATGTTTCTACCTGATGAGCCCGCTGGACGTGTGAT | 420 |
| 361 | GGCTTTTGGGAGACACCGATGTTTCTACCTGATGAGCCCGCTGGACGTGTGAT | 420 |
| 361 | GGCTTTTGGGAGACACCGATGTTTCTACCTGATGAGCCCGCTGGACGTGTGAT | 420 |
| 421 | GACCTTCAAGCAGACACCGCTGATGCTGACATGCTGCTGGGATCTGATACGGGC | 480 |
| 421 | GACCTTCAAGCAGACACCGCTGATGCTGACATGCTGCTGGGATCTGATACGGGC | 480 |
| 421 | GACCTTCAAGCAGACACCGCTGATGCTGACATGCTGCTGGGATCTGATACGGGC | 480 |
| 481 | CGCTTCTGTCGACATGAGGCTGTTGATGCTGATGCTGATGCTGCTGCTGCTGCTG | 540 |
| 481 | CGCTTCTGTCGACATGAGGCTGTTGATGCTGATGCTGATGCTGCTGCTGCTGCTG | 540 |
| 481 | CGCTTCTGTCGACATGAGGCTGTTGATGCTGATGCTGATGCTGCTGCTGCTGCTG | 540 |
| 541 | GTCACGCTCTTTCAGTGGAGAGACGCGGCTGCTGCTGCTGCTGCTGCTGCTG | 600 |
| 541 | GTCACGCTCTTTCAGTGGAGAGACGCGGCTGCTGCTGCTGCTGCTGCTGCTG | 600 |
| 541 | GTCACGCTCTTTCAGTGGAGAGACGCGGCTGCTGCTGCTGCTGCTGCTGCTG | 600 |
| 601 | GATCCCAAGATGAAT | 660 |
| 601 | GATCCCAAGATGAAT | 660 |
| 601 | GATCCCAAGATGAAT | 660 |
| 661 | CTTTGAGGTGGTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 720 |
| 661 | CTTTGAGGTGGTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 720 |
| 661 | CTTTGAGGTGGTGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 720 |
| 721 | CTTTTCAACCTGACGCTGCTTACCGCTGCTGTAAGAACCCCTTCCCTCAACCTG | 780 |
| 721 | CTTTTCAACCTGACGCTGCTTACCGCTGCTGTAAGAACCCCTTCCCTCAACCTG | 780 |
| 721 | CTTTTCAACCTGACGCTGCTTACCGCTGCTGTAAGAACCCCTTCCCTCAACCTG | 780 |
| 781 | CGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 840 |
| 781 | CGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 840 |
| 781 | CGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | 840 |
| 841 | TCTCATGATGACAGTGCATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG | 900 |

```
Db 841 TCTCATMGATGACATCGCATGTTGATGGCGAGCATGAGCAGAAACCTCAGAAAGAGAGA 900
QY 901 CCACCCCTACTATGTGATGACAGTATCTGCCAAAGCCAGCTGAGATCTCAGAACCAT 960
Db 901 CCACCCCTACTATGTGATGACAGTATCTGCCAAAGCCAGCTGAGATCTCAGAACCAT 960
QY 961 CCAGTCTTGCCCAAGGCGCTGACAGAAAGGCTACTGCTTGCGGCTATGAGAGCCTGGC 1020
Db 961 CCAGTCTTGCCCAAGGCGCTGACAGAAAGGCTACTGCTTGCGGCTATGAGAGCCTGGC 1020
QY 1021 TCGAGCCCTGTGCGCCAGACTTCCCGAATGTATGAATTCCTGGGATTTGAAATTCCTGGC 1080
Db 1021 TCGAGCCCTGTGCGCCAGACTTCCCGAATGTATGAATTCCTGGGATTTGAAATTCCTGGC 1080
QY 1081 CCATCTGAGACCTGGGAGCATATCATCCCGAGGCAAGGCAATGGGTGACCCAGCTTT 1140
Db 1081 CCATCTGAGACCTGGGAGCATATCATCCCGAGGCAAGGCAATGGGTGACCCAGCTTT 1140
QY 1141 CCACAAATGCGCAGGATGCGCTTAATGTCTCCAGGCTTGCGGCTGCTTTGGCCCTA 1200
Db 1141 CCACAAATGCGCAGGATGCGCTTAATGTCTCCAGGCTTGCGGCTGCTTTGGCCCTA 1200
QY 1201 TGAAGAAGTTTCTGCACTTCAGAAAGCCTGTGGCGATGCCATGAATTTGCTGGGCTACCG 1260
Db 1201 TGAAGAAGTTTCTGCACTTCAGAAAGCCTGTGGCGATGCCATGAATTTGCTGGGCTACCG 1260
QY 1261 CCAGCTCAGATGTGACAGAAAGCAAGAAACCTGTGCTGATCTCTGTCTACTGTGAC 1320
Db 1261 CCAGCTCAGATGTGACAGAAAGCAAGAAACCTGTGCTGATCTCTGTCTACTGTGAC 1320
QY 1321 TGTCCCTGACAAATCCACTAAGAGGTTGAGAAAGCTTTCTGCGACCTGGTGTACACC 1380
Db 1321 TGTCCCTGACAAATCCACTAAGAGGTTGAGAAAGCTTTCTGCGACCTGGTGTACACC 1380
QY 1381 TCAGTCACTTCTGTGATGCTTGGAGCTTGGAGCTTGGCTTACATCTGTGAGCTTAACATG 1440
Db 1381 TCAGTCACTTCTGTGATGCTTGGAGCTTGGAGCTTGGCTTACATCTGTGAGCTTAACATG 1440
QY 1441 TCTGTGGTATCACACTGAGTGTGATGTTGTCCACAGCTGTCTGACAGAGAAAGACTTTT 1500
Db 1441 TCTGTGGTATCACACTGAGTGTGATGTTGTCCACAGCTGTCTGACAGAGAAAGACTTTT 1500
QY 1501 GTGTCCATGCTGTGTAGAAAGAAACAGACTGGGAACTTATGTAGACAGACATCCAC 1560
Db 1501 GTGTCCATGCTGTGTAGAAAGAAACAGACTGGGAACTTATGTAGACAGACATCCAC 1560
QY 1561 CAGTGAACAGGGATTTGCTCTTTCTTTCTTTGATCTTCTGTGCTGTGGGAGACTTCAG 1620
Db 1561 CAGTGAACAGGGATTTGCTCTTTCTTTCTTTGATCTTCTGTGCTGTGGGAGACTTCAG 1620
QY 1621 AGACTTTGTGGCTGAGAGCCCTATTAACGACAGACATATCAGTGGATTTGATCCATTA 1680
Db 1621 AGACTTTGTGGCTGAGAGCCCTATTAACGACAGACATATCAGTGGATTTGATCCATTA 1680
QY 1681 ACCTCCCTGTCCACATCTTGCCCAATGGGAGATGATCTTTCACCAAGAGAGTCCAGCC 1740
Db 1681 ACCTCCCTGTCCACATCTTGCCCAATGGGAGATGATCTTTCACCAAGAGAGTCCAGCC 1740
QY 1741 ATTTTCCACAGAGATGCAAAATTCAGAGCCTTTGAGTTCCAGTGGATTCAGAGAGAA 1800
Db 1741 ATTTTCCACAGAGATGCAAAATTCAGAGCCTTTGAGTTCCAGTGGATTCAGAGAGAA 1800
QY 1801 GTGGGAAAGAGTGTGAGAGCTTATGAGATTGACATACAGCTTCGATATCAGA 1860
Db 1801 GTGGGAAAGAGTGTGAGAGCTTATGAGATTGACATACAGCTTCGATATCAGA 1860
QY 1861 AATATGAACAAATCTGTGACAAAGAGCAAGCTCTTAAGTTCACAGGAGTCCAGG 1920
Db 1861 AATATGAACAAATCTGTGACAAAGAGCAAGCTCTTAAGTTCACAGGAGTCCAGG 1920
QY 1921 TGCATTTGAATTCACCTTCCCTGTGATTTTCCATCAGATAGAAAGCTTTGACCTGTG 1980
Db 1921 TGCATTTGAATTCACCTTCCCTGTGATTTTCCATCAGATAGAAAGCTTTGACCTGTG 1980
```

```
Db 1921 TGCATTTGAATTCACCTTCCCTGTGATTTTCCATCAGATAGAAAGCTTTGACCTGTG 1980
QY 1981 AAGCTGCCATCTGTTAATTAATAATCCCAATAGCAATAGCAATAGCAATAGCAATAGCA 2032
Db 1981 AAGCTGCCATCTGTTAATTAATAATCCCAATAGCAATAGCAATAGCAATAGCAATAGCA 2032
RESULT 2
AF131235 2032 bp mRNA linear PRI 10-ANG-1999
LOCUS AF131235
DEFINITION Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA, complete cds.
ACCESSION AF131235
VERSION AF131235.1 GI:4927113
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Homo sapiens. Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2032)
Bistrup, A., Bhakta, S., Lee, J.-K., Belov, Y. Y., Gunn, M. D., Zuo, F. R., Huang, C. C., Kannagi, R., Rosen, S. D. and Hemmerich, S., The sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin J. Cell Biol. 145 (4), 899-910 (1999)
JOURNAL
MEDLINE 99264336
PUBMED 10330415
REFERENCE 2 (bases 1 to 2032)
Bistrup, A., Tangemann, K., Bhakta, S., Lee, J.-K., Belov, Y. Y., Gunn, M. D., Zuo, F. R., Huang, C. C., Kannagi, R., Rosen, S. D. and Hemmerich, S., Direct Submmission Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA
FEATURES
source 1..2032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="high endothelial"
/tissue_type="tonsil"
182..1342
/function="biosynthesis of endothelial ligands for L-selectin"
/note="HEC-GlcNAc-6-Sr"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="A033015.1"
/db_xref="GI:4927114"
/translation="MLPKRMKLLFLVSOMATLAFPMYSNITSLSMKRQPERMH VIVLSFMRSGSSFVGGOLFQHPDVEYLMHPAMVHMTFFRQSPAMRHMAVRILRAV LCMVSFDAYMEGRPQRSLSRQWENSRLCSRPADITLRODETIFRANRLLCSQF FEVEYRACRSYSYHVLKEVEFVLOSLYLLKDPDLYVMQVIOQSLIKT1QS1PKALQERYLRY GDMLDSRYVMGQHEOKLKKEDDPYVMQVIOQSLIKT1QS1PKALQERYLRY EDLARAIPVQTSRMVEFVGLFELPHQTVNHNITRKGSGDAFTNARDALNVQAM RSLPYEYRSRLQACGDAMNLLGYRHSVSEQRRLDLLDLSTVTPQRIH"
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN
Query Match 100.0%; Score 2032; DB 9; Length 2032;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 181 AATGCTACTGGCTAAAAAATGAAGCTGCTGTTTCTGTTTCCAGATGGCATCTT 240
DB 181 AATGCTACTGGCTAAAAAATGAAGCTGCTGTTTCTGTTTCCAGATGGCATCTT 240
QY 241 GGCCTATCTTCACATGTACAGCCACACATCAGCTCCCTGTCTATGAAGGCACAGCC 300
DB 241 GGCCTATCTTCACATGTACAGCCACACATCAGCTCCCTGTCTATGAAGGCACAGCC 300
QY 301 CGAGGGCATGACGCTGGTGTCTCTCTGCGGCTGCTGGCTCTTCTTGGGGGCA 360
DB 301 CGAGGGCATGACGCTGGTGTCTCTCTGCGGCTGCTGGCTCTTCTTGGGGGCA 360
QY 361 GCTTTTGGGAGACCCAGATGTTTCTTACCTGATGAGGCCCGCTGGGACGTGTGGAT 420
DB 361 GCTTTTGGGAGACCCAGATGTTTCTTACCTGATGAGGCCCGCTGGGACGTGTGGAT 420
QY 421 GACCTTCAAGACAGACCCGCTGGATGCTGCATGAGCTGTGGGATCTGATACGGGC 480
DB 421 GACCTTCAAGACAGACCCGCTGGATGCTGCATGAGCTGTGGGATCTGATACGGGC 480
QY 481 GCTCTCTTGTGACACATGAGCGTCTTGTGATGCTTACATGGAACCTGCTCCCGGAGACA 540
DB 481 GCTCTCTTGTGACACATGAGCGTCTTGTGATGCTTACATGGAACCTGCTCCCGGAGACA 540
QY 541 GTCCAGCCTCTTTCAAGTGGGAGAACAGCCGGCCCTGTCTCACCCTGCTGTGACAT 600
DB 541 GTCCAGCCTCTTTCAAGTGGGAGAACAGCCGGCCCTGTCTCACCCTGCTGTGACAT 600
QY 601 CATCCCAAGATGAATCATCCCGGGCTCAGTGCAGGCTCTGTGAGTCAAGAGCC 660
DB 601 CATCCCAAGATGAATCATCCCGGGCTCAGTGCAGGCTCTGTGAGTCAAGAGCC 660
QY 661 CTTTGAAGTGTGAGAGGCTCCGCTCTCTACAGCCAGCTGTGTCTCAAGAGAGTGG 720
DB 661 CTTTGAAGTGTGAGAGGCTCCGCTCTCTACAGCCAGCTGTGTCTCAAGAGAGTGG 720
QY 721 CTTCTCAACCGGAGTCCCTCTACCGGCTGTGAAGACCCCTCCCAACCGGAGTAT 780
DB 721 CTTCTCAACCGGAGTCCCTCTACCGGCTGTGAAGACCCCTCCCAACCGGAGTAT 780
QY 781 CGTGCACCTGGTCCGGGAGCCCGGGCGGTTCGTTCCGAGAGCAGCAAAAGGAGA 840
DB 781 CGTGCACCTGGTCCGGGAGCCCGGGCGGTTCGTTCCGAGAGCAGCAAAAGGAGA 840
QY 841 TCTCATGTAGAGTGGCATTTGTGATGGGAGCAGTACAGAGAACTCAAGAGAGGA 900
DB 841 TCTCATGTAGAGTGGCATTTGTGATGGGAGCAGTACAGAGAACTCAAGAGAGGA 900
QY 901 CCACCCCTACTATGTGATGAGGTGATCTGCCAAGCCAGCTGAGATCTACAGAGCAT 960
DB 901 CCACCCCTACTATGTGATGAGGTGATCTGCCAAGCCAGCTGAGATCTACAGAGCAT 960
QY 961 CCAGTCTCTGCCCAAGGCCCTGCAAGAACGCTACCTGCTGTGGCTATAGAGCCGCG 1020
DB 961 CCAGTCTCTGCCCAAGGCCCTGCAAGAACGCTACCTGCTGTGGCTATAGAGCCGCG 1020
QY 1021 TCGAGCCCTGTGGGCCAGACTTCCCGAATGTGAATGTGGGATTTGGAATCTTGGC 1080
DB 1021 TCGAGCCCTGTGGGCCAGACTTCCCGAATGTGAATGTGGGATTTGGAATCTTGGC 1080
QY 1081 CCACTCTCAGACCTGGGTGATTAACATCACCCGAGGCAAGGGCATGGGTACCAAGCTTT 1140
DB 1081 CCACTCTCAGACCTGGGTGATTAACATCACCCGAGGCAAGGGCATGGGTACCAAGCTTT 1140
QY 1141 CCACACCAATGCCAGGATGCCCTTATGTCTCCAGGCTGTGGGCTGTTGGCCCTA 1200
DB 1141 CCACACCAATGCCAGGATGCCCTTATGTCTCCAGGCTGTGGGCTGTTGGCCCTA 1200
QY 1201 TGAAGAAGTTTCTGACATTCAGAAAGCCTGTGGAGTGCATGAATTTGCTGGGCTACCG 1260
DB 1201 TGAAGAAGTTTCTGACATTCAGAAAGCCTGTGGAGTGCATGAATTTGCTGGGCTACCG 1260

QY 1261 CCACGTCAAGTCTGAACAAAGACAGAAACCTGTTGCTGGAATCTCTGCTACCTGGAC 1320
DB 1261 CCACGTCAAGTCTGAACAAAGACAGAAACCTGTTGCTGGAATCTCTGCTACCTGGAC 1320
QY 1321 TGTCCCTGAGCAATTCACCTAGAGGGTTGAGAGGCTTGTCTGCTACCTGCTGACGC 1380
DB 1321 TGTCCCTGAGCAATTCACCTAGAGGGTTGAGAGGCTTGTCTGCTACCTGCTGACGC 1380
QY 1381 TCAAGTCACTTCTCTGAATGCTTCTGAGCCTGACACTCTCTGAGCTTAACTACATG 1440
DB 1381 TCAAGTCACTTCTCTGAATGCTTCTGAGCCTGACACTCTCTGAGCTTAACTACATG 1440
QY 1441 TCTGTGGATTCACACTGAGTGTGATGTGTCCACAGCTGCTCAAGACAGAGACTTTT 1500
DB 1441 TCTGTGGATTCACACTGAGTGTGATGTGTCCACAGCTGCTCAAGACAGAGACTTTT 1500
QY 1501 GTGTCCATGCTGTGTGTGAAGAACAGCTGGGGAACCTTATGTAGACGACATCCAC 1560
DB 1501 GTGTCCATGCTGTGTGTGAAGAACAGCTGGGGAACCTTATGTAGACGACATCCAC 1560
QY 1561 CAGTGAACAGGGTATGCTCTTCTTCTTCTTCTTCTTCTTCTGCTGCTGGGAGACTTCAG 1620
DB 1561 CAGTGAACAGGGTATGCTCTTCTTCTTCTTCTTCTTCTTCTGCTGCTGGGAGACTTCAG 1620
QY 1621 AGACTTTGTGGCTGGAGGCTTATTAAGCAGACAGATATAGTGAATGTATCCATTA 1680
DB 1621 AGACTTTGTGGCTGGAGGCTTATTAAGCAGACAGATATAGTGAATGTATCCATTA 1680
QY 1681 ACCTCCCTGCTCCACATCTTGCCCAATGAGGGAATGATCTTTCACCAAGAGCTCACAGC 1740
DB 1681 ACCTCCCTGCTCCACATCTTGCCCAATGAGGGAATGATCTTTCACCAAGAGCTCACAGC 1740
QY 1741 ATTTTCCACAGATGCAAAATTTGAGCCCTTGGAGTTTCCAGTGGATTTCAAGAGAGA 1800
DB 1741 ATTTTCCACAGATGCAAAATTTGAGCCCTTGGAGTTTCCAGTGGATTTCAAGAGAGA 1800
QY 1801 GTGGAAACAAGTGTGATGCTTACTTATGAGCTTGACATTCACAGTACGCTATTCAGA 1860
DB 1801 GTGGAAACAAGTGTGATGCTTACTTATGAGCTTGACATTCACAGTACGCTATTCAGA 1860
QY 1861 AATATGAACAAATCTGTGACACAAAGACAGCTTAACTTCAAGAGGCTCTGGGC 1920
DB 1861 AATATGAACAAATCTGTGACACAAAGACAGCTTAACTTCAAGAGGCTCTGGGC 1920
QY 1921 TGCATTGAATACACTTCCCTGCTGATTTTCCATGCAATGAAGACTTGTGACCTGTG 1980
DB 1921 TGCATTGAATACACTTCCCTGCTGATTTTCCATGCAATGAAGACTTGTGACCTGTG 1980
QY 1981 AAGCTGCATCTGTTAATCTAAATTTCCCAATTAAGAAAAA 2032
DB 1981 AAGCTGCATCTGTTAATCTAAATTTCCCAATTAAGAAAAA 2032

RESULT 3
AK026635
LOCUS AK026635 2011 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ22982 fls, clone KAT11454, highly similar to AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA.
ACCESSION AK026635
VERSION AK026635.1 GI:10439531
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens signet-ring cell carcinoma cell line; KATO III CDNA to mRNA, clone lib: KAT clone: KAT11454.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1111)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
TITLE Shibahara,T., Tanaka,T. and Nakamura,Y.
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT MEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Source Location/Qualifiers
1..2011
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="KAT11454"
/cell_line="KATO III"
/cell_type="signet-ring cell carcinoma"
/clone_lib="KAT"
/note="cloning vector pME18SFL3"
1..2011
/note="highly similar to AF131235 Homo sapiens
N-acetylglucosamine 6-O-sulfotransferase mRNA"
BASE COUNT 482 a 541 c 475 g 513 t
ORIGIN
Query Match 91.3%; Score 1856; DB 9; Length 2011;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1906; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 111 GTTACTTTACAGCTTCTCTGGAGACGAGTCTTTCTCAGCCGCTCTTGAAGCTCTCC 170
Db 1 GTTACTTTACAGCTTCTCTGGAGACGAGTCTTTCTCAGCCGCTCTTGAAGCTCTCC 60
QY 171 ACTTCAGACAATGCTACTGCTCTTAAATAATGAAGTCCGTGTTCTGTCTCCAGA 230
Db 61 ACTTCAGACAATGCTACTGCTCTTAAATAATGAAGTCCGTGTTCTGTCTCCAGA 120
QY 231 TGGCATTCTGGCTCTATTCTTCCATGTACAGCCACAATCAGCTCCCTGTATGA 290
Db 121 TGGCATTCTGGCTCTATTCTTCCATGTACAGCCACAATCAGCTCCCTGTATGA 180
QY 291 AGGCACAGCCGAGCGCATGAGTGTGTCTCTCTGGCGCTGTGGCTCTTCTT 350
Db 181 AGGCACAGCCGAGCGCATGAGTGTGTCTCTCTGGCGCTGTGGCTCTTCTT 240
QY 351 TTGTGGGGCAGCTTTTGGCAGACCCAGATGTTTCTACTGTAGTGAAGCCGCGTGGC 410
Db 241 TTGTGGGGCAGCTTTTGGCAGACCCAGATGTTTCTACTGTAGTGAAGCCGCGTGGC 300
QY 411 ACCTGTGGATGACCTTTAAGCAGACGCCCTGGATGCTGTGCATATGCTGTGGGATC 470
Db 301 ACCTGTGGATGACCTTTAAGCAGACGCCCTGGATGCTGTGCATATGCTGTGGGATC 350
QY 471 TGATACGGGCGCTTCTTGTGGACATGAGCGTCTTGTGCTCATGGAAGCTGTGC 530
Db 361 TGATACGGGCGCTTCTTGTGGACATGAGCGCTTGTGCTCATGGAAGCTGTGC 420
QY 531 CCCGGAACAGTCCAGCCTTTTCAGTGGAGAAACGCCGGCGCTGTGTTCTGCACCTG 590
Db 421 CCCGGAACAGTCCAGCCTTTTCAGTGGAGAAACGCCGGCGCTGTGTTCTGCACCTG 480
QY 591 CCTGTGACATCATCCCAAGATGAATCATCCCGCGGCTCATGTGAGGCTCCTGTGCA 650
Db 481 CCTGTGACATCATCCCAAGATGAATCATCCCGCGGCTCATGTGAGGCTCCTGTGCA 540
QY 651 GTCAACAGCCCTTTGAGTGTGGAGAGGCTGTCCGCTCTCAAGCCACGTGTGTCTCA 710
|||||

Db 541 GTCAACAGCCCTTTGAGTGTGGAGAGGCTGCCGCTCTTACAGCCACGTGTGTCTCA 600
QY 711 AAGAGGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAAGACCCCTCCCTCA 770
Db 601 AAGAGGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAAGACCCCTCCCTCA 660
QY 771 ACCTGCATTCGTGCACCTGTGTCCGGACCCCGGGCGCTGTTCCTGTTCCGAGAACCA 830
Db 661 ACCTGCATTCGTGCACCTGTGTCCGGACCCCGGGCGCTGTTCCTGTTCCGAGAACCA 720
QY 831 CAAAGGACATTCATGATGATGACAGTGCATGATGTGATGGAGCAGATGACAGCAACTCA 890
Db 721 CAAAGGACATTCATGATGATGACAGTGCATGATGTGATGGAGCAGATGACAGCAACTCA 780
QY 891 AGAAGGAGACCAACCTTACTATGTGATGCAAGTCAATGTGCCAAAGCCAGCTGGAGATCT 950
Db 781 AGAAGGAGACCAACCTTACTATGTGATGCAAGTCAATGTGCCAAAGCCAGCTGGAGATCT 840
QY 951 ACAAGACATCCAGTCTTGCCCAAGGCCCTGAGAAAGCTACCTGCTGTGGCCTATG 1010
Db 841 ACAAGACATCCAGTCTTGCCCAAGGCCCTGAGAAAGCTACCTGCTGTGGCCTATG 900
QY 1011 AGGACCTGGCTGAGCCCTGTGGCCACAGCTTCCGAAATGATGAATTCGTGGATTGG 1070
Db 901 AGGACCTGGCTGAGCCCTGTGGCCACAGCTTCCGAAATGATGAATTCGTGGATTGG 960
QY 1071 AATCTTGGCCCATCTTTCAGACTGGGTGCATATCATCACCAGGACAGGACATGGGTG 1130
Db 961 AATCTTGGCCCATCTTTCAGACTGGGTGCATATCATCACCAGGACAGGACATGGGTG 1020
QY 1131 ACCAGCCTTTCACACAAATGCGAGGATGCCCTTAATGTGCCAGGCTTGGCCCTGAT 1190
Db 1021 ACCAGCCTTTCACACAAATGCGAGGATGCCCTTAATGTGCCAGGCTTGGCCCTGAT 1080
QY 1191 CTTTGGCCCATGAAAAGTTTCTCGACTTCAGAAAGCCGTGGCGATGCGAATTTGTC 1250
Db 1081 CTTTGGCCCATGAAAAGTTTCTCGACTTCAGAAAGCCGTGGCGATGCGAATTTGTC 1140
QY 1251 TGGGCTACCGCCACGTCATCTGGAACAGAAAGAACTGTGGGGAATCTTCTGT 1310
Db 1141 TGGGCTACCGCCACGTCATCTGGAACAGAAAGAACTGTGGGGAATCTTCTGT 1200
QY 1311 CTACTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTGTCTGCACCT 1370
Db 1201 CTACTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTGTCTGCACCT 1260
QY 1371 GGTGTGAGCCTGAGTCACTTCTCTGAATGCTTGTGAGCCTTGCCTACATCTCTAGCCT 1430
Db 1261 GGTGTGAGCCTGAGTCACTTCTCTGAATGCTTGTGAGCCTTGCCTACATCTCTAGCCT 1320
QY 1431 TAACATCATGTCTGTGGATACACTGAGTGTGATGTGTCCACACAGTGTCTCAAGAG 1490
Db 1321 TAACATCATGTCTGTGGATACACTGAGTGTGATGTGTGTCCACACAGTGTCTCAAGAG 1380
QY 1491 AAGGACTTTTGTGTCATCTGTGTGTAGAAAACAGATGGGGAACTTATGTAGCAG 1550
Db 1381 AAGGACTTTTGTGTCATCTGTGTGTGTAGAAAACAGATGGGGAACTTATGTAGCAG 1440
QY 1551 CACATCCCAACAGTGAAGAGGATATGTCTCTTCTTGTGATCTCTCTGCTGGG 1610
Db 1441 CACATCCCAACAGTGAAGAGGATATGTCTCTTCTTCTTGTGATCTCTCTGCTGGG 1500
QY 1611 CAGACTTCAGAGACTTTGTGGCCTGGAGGCTATTATAGCAGACAGATATCATGTGAAT 1670
Db 1501 CAGACTTCAGAGACTTTGTGGCCTGGAGGCTATTATAGCAGACAGATATCATGTGAAT 1560
QY 1671 TGATCATTAACCTCCCTCTCCACATCTTGCCCAATGGGGAATGATCTTTACCAAGA 1730
Db 1561 TGATCATTAACCTCCCTCTCCACATCTTGCCCAATGGGGAATGATCTTTACCAAGA 1620
QY 1731 GCTCACACAGCAATTTTCACAGAGATGCAAAATTCAGACCTTGAAGTTCACAGTGAATTC 1790
Db 1621 GCTCACACAGCAATTTTCACAGAGATGCAAAATTCAGACCTTGAAGTTCACAGTGAATTC 1680
|||||

| | | | | | |
|------------------------|--------|---|--------------|---------------|----------------|
| Query Match: | | 88.9% | Score 1806: | DB 9 | Length 183228: |
| Best Local Similarity: | | 99.98% | Pred. No. 0: | | |
| Matches 1856: | | Conservative | 0: | Mismatches 1: | Indels 0: |
| | | | | | Gaps |
| QY | 161 | AAGGCTTCCACTTCAGCCACATGCTACTGCTCTAAAAAATGAAGCTCTCGCTGTTTCG | 220 | | |
| Db | 142584 | AAGGCTTCCACTTCAGCCACATGCTACTGCTCTAAAAAATGAAGCTCTCGCTGTTTCG | 142644 | | |
| QY | 221 | GTTTCCCAAGATGGCCATTTGGCTCTATTTCTCCACATGTACACCCCAACATCAGTCC | 280 | | |
| Db | 142644 | GTTTCCCAAGATGGCCATTTGGCTCTATTTCTCCACATGTACACCCCAACATCAGTCC | 142703 | | |
| QY | 281 | CTGTATTGAAGGACACACCCCGACGGCATGCAAGCTGTGGTTCTGTTCTCGGCGCTCT | 340 | | |
| Db | 142704 | CTGTATTGAAGGACACACCCCGACGGCATGCAAGCTGTGGTTCTGTTCTCGGCGCTCT | 142765 | | |
| QY | 341 | GAGCTCTTTTGTGTGGGACACTTTTGGGACGACACCAGATGTTTCTACTGTATGGAG | 400 | | |
| Db | 142764 | GAGCTCTTTTGTGTGGGACACTTTTGGGACGACACCAGATGTTTCTACTGTATGGAG | 142823 | | |
| QY | 401 | CCCGCTGGACAGTGTGATACCTTCAGAGAGACACCGGCTGGATGTGCACATGGCT | 460 | | |
| Db | 142824 | CCCGCTGGACAGTGTGATACCTTCAGAGAGACACCGGCTGGATGTGCACATGGCT | 142883 | | |
| QY | 461 | GTGCGGATCTGATACGGGCGCTTCTTGTGGACATGAGCGTCTTGTATGCCCTACATG | 520 | | |
| Db | 142884 | GTGCGGATCTGATACGGGCGCTTCTTGTGGACATGAGCGCTTGTATGCCCTACATG | 142943 | | |
| QY | 521 | GAACTGTCTCCCGGAGACATCCAGGCTCTTACGTGGGAGAAACACCGGGCGCTGCT | 580 | | |
| Db | 142944 | GAACTGTCTCCCGGAGACATCCAGGCTCTTACGTGGGAGAAACACCGGGCGCTGCT | 143003 | | |
| QY | 581 | TCTGCACCTGCCGTGATCATCTCCCAACATATAATCATCCCGGCGGTCACTGAGG | 640 | | |
| Db | 143004 | TCTGCACCTGCCGTGATCATCTCCCAACATATAATCATCCCGGCGGTCACTGAGG | 143063 | | |
| QY | 641 | CTCTCTGTCACTCAACACCCCTTTGAGGTGTTGAGAAAGGCTTCCGCTCTACAGCCAC | 700 | | |
| Db | 143064 | CTCTCTGTCACTCAACACCCCTTTGAGGTGTTGAGAAAGGCTTCCGCTCTACAGCCAC | 143123 | | |
| QY | 701 | GTGGTGTCAAGGAGGTGGCTTCTTCAACTGTGAGTCCCTCTACCCGCTGTGCAAAAGAC | 760 | | |
| Db | 143124 | GTGGTGTCAAGGAGGTGGCTTCTTCAACTGTGAGTCCCTCTACCCGCTGTGCAAAAGAC | 143183 | | |
| QY | 761 | CCCTCTCCCTCAACCTGTGCATATGTGTGACCTGGTCCGGGACCCCGGGCGGTTCGTTCC | 820 | | |
| Db | 143184 | CCCTCTCCCTCAACCTGTGCATATGTGTGACCTGGTCCGGGACCCCGGGCGGTTCGTTCC | 143243 | | |
| QY | 821 | CGAGAACCCACAAGAGGAGTCTCATGTATGATACAGTGCATTTGATGGGGACATGAG | 880 | | |
| Db | 143244 | CGAGAACCCACAAGAGGAGTCTCATGTATGATACAGTGCATTTGATGGGGACATGAG | 143303 | | |
| QY | 881 | CAGAAACTCAAGAGAGAGGACCAACCTCTACTATGTATGACAGTTCATCTGCAAAAGCCAG | 940 | | |
| Db | 143304 | CAGAAACTCAAGAGAGAGGACCAACCTCTACTATGTATGACAGTTCATCTGCAAAAGCCAG | 143363 | | |
| QY | 941 | CTGCGATCTCAAGACCAATCCAGTCTGTGCCAAAGGCCGTGCAGAGAAAGCTTACCTCTT | 1000 | | |
| Db | 143364 | CTGCGATCTCAAGACCAATCCAGTCTGTGCCAAAGGCCGTGCAGAGAAAGCTTACCTCTT | 143423 | | |
| QY | 1001 | GTGCGCTATGAGGAGCTGGCTCGAGCCCTGTGGCCCAAGCTTCCGAAATGTATGAATTC | 1060 | | |
| Db | 143424 | GTGCGCTATGAGGAGCTGGCTCGAGCCCTGTGGCCCAAGCTTCCGAAATGTATGAATTC | 143483 | | |
| QY | 1061 | GTGGATTTGCAATTTCTGCCCCCATCTTTCAGACCTGGGGGTCAATACATACCCCAAGGCAAG | 1120 | | |
| Db | 143484 | GTGGATTTGCAATTTCTGCCCCCATCTTTCAGACCTGGGGGTCAATACATACCCCAAGGCAAG | 143543 | | |
| QY | 1121 | GGCATGGGTGACCAAGCTTTTCACACAAATCCAGAGATGCCCTTAATGTCTCCAGGCT | 1180 | | |
| Db | 143544 | GGCATGGGTGACCAAGCTTTTCACACAAATCCAGAGATGCCCTTAATGTCTCCAGGCT | 143603 | | |
| QY | 1181 | TGGCGCTGGCTCTTGGCCCTATGAAAGGTTTCTGACTTCAAGAAAGCCTTGGCGATGCC | 1240 | | |


```

Db 143604 TGGCGCTGGTCTTTGCCCTATGAAAGGTTCTCGACTTCAGAAAGCCTGTGGCATGCC 143663
OY 1241 ATGAATTTGCTGGGGCTACCGGCACGCTGACATCTGAAACAAGAACAGAACTGTGTG 1300
Db 143664 ATGAATTTGCTGGGGCTACCGGCACGCTGACATCTGAAACAAGAACAGAACTGTGTG 143723
OY 1301 GATCTTCTGTCTACCTGGAGTGTCCCTAGCAAAATCCACTAAGAGGGTTGAAAGGCTTT 1360
Db 143724 GATCTTCTGTCTACCTGGAGTGTCCCTAGCAAAATCCACTAAGAGGGTTGAAAGGCTTT 143783
OY 1361 GCTGCCACCTGGTGTACAGCTTCAGCTTTCTCTGAAATGCTTCTGAGCCTTGCCCTACAT 1420
Db 143784 GCTGCCACCTGGTGTACAGCTTCAGCTTTCTCTGAAATGCTTCTGAGCCTTGCCCTACAT 143843
OY 1421 CTCTAGGCTTACTACATGCTGTGTGGTATCACACTAGTGTGATGTGTCCACAGCT 1480
Db 143844 CTCTAGGCTTACTACATGCTGTGTGGTATCACACTAGTGTGATGTGTGTCCACAGCT 143903
OY 1481 GCTCAGAGAGAGAGACTTTTGTGTCCATGCTTGTGTCTAGAAAACAGACTGGGAGACCTT 1540
Db 143904 GCTCAGAGAGAGAGACTTTTGTGTCCATGCTTGTGTCTAGAAAACAGACTGGGAGACCTT 143963
OY 1541 ATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1600
Db 143964 ATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144023
OY 1601 CCTGTCTGGGAGAGACTTCAGAGACTTTGTGGCCTGGAGCCCTATTAAACAGACAGAGTA 1660
Db 144024 CCTGTCTGGGAGAGACTTCAGAGACTTTGTGGCCTGGAGCCCTATTAAACAGACAGAGTA 144083
OY 1661 TCAGTGGATTTGATCCATAAAGCTCCCTGTCCACTCTTCCCAATGGGAGATGATCTT 1720
Db 144084 TCAGTGGATTTGATCCATAAAGCTCCCTGTCCACTCTTCCCAATGGGAGATGATCTT 144143
OY 1721 TCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1780
Db 144144 TCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144203
OY 1781 CAGTGGATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1840
Db 144204 CAGTGGATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144263
OY 1841 CACAGCTTCGAGTATCAAGAAATATGAACAAATCTGTGACAAAGAGAGAGAGAGAGAGAG 1900
Db 144264 CACAGCTTCGAGTATCAAGAAATATGAACAAATCTGTGACAAAGAGAGAGAGAGAGAGAG 144323
OY 1901 AGTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1960
Db 144324 AGTTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144383
OY 1961 ATAGAGAGCTTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGTGAAG 2020
Db 144384 ATAGAGAGCTTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGTGAAGCTGTGAAG 144440

```

```

JOURNAL 6-O-sulfoltransferase gene family
MEDLINE Glycobiology 11 (1), 75-87 (2001)
PUBMED 21096027
REFERENCE 1181564
AUTHORS 2 (bases 1 to 1992)
Hemmerich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddle, N.R. and
Rosen, S.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES
source
1..1992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1-q23.2"
/cell_type="high endothelial"
/tissue_type="tonsil"
1..1992
/gene="GST3"
122..1282
/gene="GST3"
/notes="similar to Homo sapiens sequence encoded by Genbank
Accession Number AF131235"
/codon_start=1
/product="L-selectin ligand sulfoltransferase GST-3"
/db_xref="GI:12060808"
/translation="MLPRKMKLLFLVSGMAALLFFHMSHNTSISMKOPRHH
VILSSNRGSSPVGQLEQHPDVEYILMPAHVWTFEOSTAMLHNAVRLITAVF
LCMSVFDVIMGEPRKQSLFQWENSRALCSAPCDITPDEILIPRAICRLCSQP
FEVEKACRSYSHVILKEVRFNLSVLELDPISLNIIVLVDPRVFRSREIRK
GDIMDSRIVMGHDEKLEKEDQPYVMVQICQSLQSLPKALOERYLLIVRY
EDLARAPVQTSMTSEFVLEFLEPLHLOTWVNTTRGKGDHAFPTNARDALNVQAW
RMSLPYEKYSRLQKAGDAMNLGVRHVSDEQRLLDLISTVTPQJH"
BASE COUNT 482 a 540 c 489 g 481 t
ORIGIN
Query Match 86.4% Score 1756; DB 9; Length 1992;
Best local Similarity 99.9%; Pred. No. 0;
Matches 1856; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 160 CAGAGCTTCACACTTCAGCAATGCTACTGCTTAATAAATAGACTCTGCTTTCT 219
Db 100 CAGAGCTTCACACTTCAGCAATGCTACTGCTTAATAAATAGACTCTGCTTTCT 159
OY 220 GGTTCACAGATGGCATTCTGGCTATCTTCCACANTGTACAGCCACACATCAGCTC 279
Db 160 GGTTCACAGATGGCATTCTGGCTATCTTCCACANTGTACAGCCACACATCAGCTC 219
OY 280 CTTGTCTATAGAGCAGAGCCCGAGCGCATGCACTGCTGCTCTTCTCGGGGCTC 339
Db 220 CTTGTCTATAGAGCAGAGCCCGAGCGCATGCACTGCTGCTCTTCTCGGGGCTC 279
OY 340 TGGCTCTTCTTGTGTGGGAGAGCTTTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 399
Db 280 TGGCTCTTCTTGTGTGGGAGAGCTTTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 339
OY 400 GCCCGCTGGCAGAGTGTGATGACCTTCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
Db 340 GCCCGCTGGCAGAGTGTGATGACCTTCAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 399
OY 460 TGTGCGGAGATCTGATAGAGGCGGCTTCTGTGCGAGATCAGGCTCTTTGATGCTACAT 519
Db 400 TGTGCGGAGATCTGATAGAGGCGGCTTCTGTGCGAGATCAGGCTCTTTGATGCTACAT 459
OY 520 GGAACCTGTGTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 579
Db 460 GGAACCTGTGTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519
OY 580 TTTCGACCTGCTGTGACATCCACAGAGATGAATATGATCCCGGGGCTACATGACAG 639
Db 520 TTTCGACCTGCTGTGACATCCACAGAGATGAATATGATCCCGGGGCTACATGACAG 579

```

| | | | |
|----|------|---|------|
| QY | 640 | GCTCCTGTGAGTCAACAGCCCTTTGAGTGTGAGAGAGAGCCCTGCGGCTCTACACCA | 699 |
| Db | 580 | GCTCCTGTGAGTCAACAGCCCTTTGAGTGTGAGAGAGAGCCCTGCGGCTCTACACCA | 639 |
| QY | 700 | CGTGTGCTCAAGAGGCGCTTCTTCAACCTGACAGTCCCTTCAACCGGTGTGAAGA | 759 |
| Db | 640 | CGTGTGCTCAAGAGGCGCTTCTTCAACCTGACAGTCCCTTCAACCGGTGTGAAGA | 699 |
| QY | 760 | CCCCCTCCACCTGCATATGTCGACGTGTCCGGAGCCCGGGCGGTGTCCGCTTC | 819 |
| Db | 700 | CCCCCTCCACCTGCATATGTCGACGTGTCCGGAGCCCGGGCGGTGTCCGCTTC | 759 |
| QY | 820 | CCGAGAACGCAAGAGGAGATCATGATGACAGTCGATGTGATGAGGAGCAGTGA | 879 |
| Db | 760 | CCGAGAACGCAAGAGGAGATCATGATGACAGTCGATGTGATGAGGAGCAGTGA | 819 |
| QY | 880 | GCAGAACTCAAGAGGAGGACCAACCCCTACTATGTATGATGAGTCACTGTGCCAAGCA | 939 |
| Db | 820 | GCAGAACTCAAGAGGAGGACCAACCCCTACTATGTATGATGAGTCACTGTGCCAAGCA | 879 |
| QY | 940 | GCTGAGATCTACAAAGACATCCAGTCTGTCGCCAGGCGCTGAGAGAGCCTACCTGCT | 999 |
| Db | 880 | GCTGAGATCTACAAAGACATCCAGTCTGTCGCCAGGCGCTGAGAGAGCCTACCTGCT | 939 |
| QY | 1000 | TGTGGCTATGAGGAGCCTGCTCGAGCCCTGTGTGGCCAGACTTCCGAATGTATGAAT | 1059 |
| Db | 940 | TGTGGCTATGAGGAGCCTGCTCGAGCCCTGTGTGGCCAGACTTCCGAATGTATGAAT | 999 |
| QY | 1060 | CGTGGGATTTGAATTTCTTGGCCCATCTTCAGACTGGGTGATATACATCACCCAGCA | 1119 |
| Db | 1000 | CGTGGGATTTGAATTTCTTGGCCCATCTTCAGACTGGGTGATATACATCACCCAGCA | 1059 |
| QY | 1120 | GGGAGTGGGTACACAGCGTTTCCACACAATGCGCAGGAGATCCCTTATGTCTCCAGGC | 1179 |
| Db | 1060 | GGGAGTGGGTACACAGCGTTTCCACACAATGCGCAGGAGATCCCTTATGTCTCCAGGC | 1119 |
| QY | 1180 | TTGGCGCTGTCTTTGGCTTATGAAGAGTTTCTCGACTTACGAAGCCTGTGGCAGTGC | 1239 |
| Db | 1120 | TTGGCGCTGTCTTTGGCTTATGAAGAGTTTCTCGACTTACGAAGCCTGTGGCAGTGC | 1179 |
| QY | 1240 | CATGATTTGCTGGGCTACCGCCAGTCAGATCTGAACAAGACAGAGAAACCTGTGCT | 1299 |
| Db | 1180 | CATGATTTGCTGGGCTACCGCCAGTCAGATCTGAACAAGACAGAGAAACCTGTGCT | 1239 |
| QY | 1300 | GGATTTGCTGTACCTGTGAGCTGTCCCTGAGCAATCCACATGAAGAGGTTGAGAGGCTT | 1359 |
| Db | 1240 | GGATTTGCTGTACCTGTGAGCTGTCCCTGAGCAATCCACATGAAGAGGTTGAGAGGCTT | 1299 |
| QY | 1360 | TGCTGCCACCTGGTGTGAGCTCACTGCTCTGTGAATGCTCTGAGCCTTGCTTACA | 1419 |
| Db | 1300 | TGCTGCCACCTGGTGTGAGCTCACTGCTCTGTGAATGCTCTGAGCCTTGCTTACA | 1359 |
| QY | 1420 | TCTGTGAGCCTTAACATACATGTGTGTGATACACAGTGAAGTGTGTGTCCACAG | 1479 |
| Db | 1360 | TCTGTGAGCCTTAACATACATGTGTGTGATACACAGTGAAGTGTGTGTGTCCACAG | 1419 |
| QY | 1480 | TGCTTACACAAAGAGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1539 |
| Db | 1420 | TGCTTACACAAAGAGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1479 |
| QY | 1540 | TATGTGAGCAGCAGATCCACAGTGAAGAGGATGTGCTCTCTCTCTCTCTCTCTCT | 1599 |
| Db | 1480 | TATGTGAGCAGCAGATCCACAGTGAAGAGGATGTGCTCTCTCTCTCTCTCTCTCT | 1539 |
| QY | 1600 | TCTGTGTGAGCAGCTTACAGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1659 |
| Db | 1540 | TCTGTGTGAGCAGCTTACAGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT | 1599 |
| QY | 1660 | ATGATGTGAATTTGAATCCATTAACCTCCCTGTCCACATCTTGTCCAAAGGGGAATGATCT | 1719 |
| Db | 1600 | ATGATGTGAATTTGAATCCATTAACCTCCCTGTCCACATCTTGTCCAAAGGGGAATGATCT | 1659 |
| QY | 1720 | TTACCAAGAGCTCCACGATTTTCCAGAGAGATGGAATTTGAGCCCTTGAGGTTTC | 1779 |
| Db | 1660 | TTACCAAGAGCTCCACGATTTTCCAGAGAGATGGAATTTGAGCCCTTGAGGTTTC | 1719 |
| QY | 1780 | CCAGTGGATTTCAAGAGGAGTGGGACAGTGTGATGCTCTACTTATGAGCTTGACCA | 1839 |
| Db | 1720 | CCAGTGGATTTCAAGAGGAGTGGGACAGTGTGATGCTCTACTTATGAGCTTGACCA | 1779 |
| QY | 1840 | TCACAGCTATGCGTATTCAGAAATATGAACAATCTGTGCACAAAGAGAGAGCTCTT | 1899 |
| Db | 1780 | TCACAGCTATGCGTATTCAGAAATATGAACAATCTGTGCACAAAGAGAGAGCTCTT | 1839 |
| QY | 1900 | AACTTACAGAGGCTGCTGGCTGCATTTGATATACACTTCCCTGTGATTTCCATCA | 1959 |
| Db | 1840 | AACTTACAGAGGCTGCTGGCTGCATTTGATATACACTTCCCTGTGATTTCCATCA | 1899 |
| QY | 1960 | CATAGAGACTTTGAGCTGTGAGAGCTGCCATCTCTTAATACATTAATTTCCCAATAG | 2017 |
| Db | 1900 | CATAGAGACTTTGAGCTGTGAGAGCTGCCATCTCTTAATACATTAATTTCCCAATAG | 1957 |

| | | | |
|----|------|---|------|
| QY | 1720 | TTACCAAGAGCTCCACGATTTTCCAGAGAGATGGAATTTGAGCCCTTGAGGTTTC | 1779 |
| Db | 1660 | TTACCAAGAGCTCCACGATTTTCCAGAGAGATGGAATTTGAGCCCTTGAGGTTTC | 1719 |
| QY | 1780 | CCAGTGGATTTCAAGAGGAGTGGGACAGTGTGATGCTCTACTTATGAGCTTGACCA | 1839 |
| Db | 1720 | CCAGTGGATTTCAAGAGGAGTGGGACAGTGTGATGCTCTACTTATGAGCTTGACCA | 1779 |
| QY | 1840 | TCACAGCTATGCGTATTCAGAAATATGAACAATCTGTGCACAAAGAGAGAGCTCTT | 1899 |
| Db | 1780 | TCACAGCTATGCGTATTCAGAAATATGAACAATCTGTGCACAAAGAGAGAGCTCTT | 1839 |
| QY | 1900 | AACTTACAGAGGCTGCTGGCTGCATTTGATATACACTTCCCTGTGATTTCCATCA | 1959 |
| Db | 1840 | AACTTACAGAGGCTGCTGGCTGCATTTGATATACACTTCCCTGTGATTTCCATCA | 1899 |
| QY | 1960 | CATAGAGACTTTGAGCTGTGAGAGCTGCCATCTCTTAATACATTAATTTCCCAATAG | 2017 |
| Db | 1900 | CATAGAGACTTTGAGCTGTGAGAGCTGCCATCTCTTAATACATTAATTTCCCAATAG | 1957 |

| | | | |
|----|-----|--|-----|
| QY | 161 | AAAGTTCACACTTCAGCAATGCTACTGCTTAAATAATGAGTCTGCTGTTCTG | 220 |
| Db | 72 | AAAGTTCACACTTCAGCAATGCTACTGCTTAAATAATGAGTCTGCTGTTCTG | 131 |

| | | | |
|----|------|--|------|
| QY | 221 | GTTCCTCCAGATGGCACTTCGGCTCTATCTTCACATGTACAGCCAACTCAGCTCC | 280 |
| Db | 132 | GTTCCTCCAGATGGCACTTCGGCTCTATCTTCACATGTACAGCCAACTCAGCTCC | 191 |
| QY | 281 | CTGTCTATGAAGGCAAGCCCGAAGCGCATCCAGCTGCTGTCTCTCTCGCGCTCT | 340 |
| Db | 192 | CTGTCTATGAAGGCAAGCCCGAAGCGCATCCAGCTGCTGTCTCTCTCGCGCTCT | 251 |
| QY | 341 | GGCTCTCTTTGTGGGGCAGCTTTTGGGCACACCCAGATGTTTTTCTACTGATGAG | 400 |
| Db | 252 | GGCTCTCTTTGTGGGGCAGCTTTTGGGCACACCCAGATGTTTTTCTACTGATGAG | 311 |
| QY | 401 | CCCGCTGGGACAGTGTGGATGGAGACCTTCAAGCAGACAGCCGCTGGATGCTGCACATGGCT | 460 |
| Db | 312 | CCCGCTGGGACAGTGTGGATGGAGACCTTCAAGCAGACAGCCGCTGGATGCTGCACATGGCT | 371 |
| QY | 461 | GTGGGGGATCTGATAGGGGCGCTCTCTGTGTGGACATAGCGCTCTTGATGCTTACATG | 520 |
| Db | 372 | GTGGGGGATCTGATAGGGGCGCTCTCTGTGTGGACATAGCGCTCTTGATGCTTACATG | 431 |
| QY | 521 | GAACCTGATCCCCGGAGAGCAAGTCCAGCCTCTTTCACTGGGAACAAGCCGAGCCCTGTGT | 580 |
| Db | 432 | GAACCTGATCCCCGGAGAGCAAGTCCAGCCTCTTTCACTGGGAACAAGCCGAGCCCTGTGT | 491 |
| QY | 581 | TCTGCACTGCTGTGTGACATCATTCCTCAACAAGATGAATATATCCCCGGGCTCATGCAAG | 640 |
| Db | 492 | TCTGCACTGCTGTGTGACATCATTCCTCAACAAGATGAATATATCCCCGGGCTCATGCAAG | 551 |
| QY | 641 | CTCTGTGCAGTCAACAGCCCTTTGAGGTGTGAGCAAGGCTGGCGCTCCTACAGCCAC | 700 |
| Db | 552 | CTCTGTGCAGTCAACAGCCCTTTGAGGTGTGAGCAAGGCTGGCGCTCCTACAGCCAC | 611 |
| QY | 701 | GTGTGCTCAAGGAGGTGGCTTTCTTCAACTGCTCAGTCCCTCTACCCGCTGCTGAAGAC | 760 |
| Db | 612 | GTGTGCTCAAGGAGGTGGCTTTCTTCAACTGCTCAGTCCCTCTACCCGCTGCTGAAGAC | 671 |
| QY | 761 | CCCTCCCTCAACCGCATATGTGTGACACTGTGTCGGGACCCCGGCGGTGTTCGGTACC | 820 |
| Db | 672 | CCCTCCCTCAACCGCATATGTGTGACACTGTGTCGGGACCCCGGCGGTGTTCGGTACC | 731 |
| QY | 821 | CGAAGACGCACAAGGGGAGATCTCATGATTAGACAGTCGANTGTATGGGGGAGCATGAG | 880 |
| Db | 732 | CGAAGACGCACAAGGGGAGATCTCATGATTAGACAGTCGANTGTATGGGGGAGCATGAG | 791 |
| QY | 881 | CAGAACTCAAGAAGGAGGAGCAACCTCTACTATGTATGCTAGAGTATCTGCCAAAGCCAG | 940 |
| Db | 792 | CAGAACTCAAGAAGGAGGAGCAACCTCTACTATGTATGCTAGAGTATCTGCCAAAGCCAG | 851 |
| QY | 941 | CTGAGATCTACAGAGACATCAACACTGCTTGGCCCAAGCCCTCAGAGAGCTACCTGCTT | 1000 |
| Db | 852 | CTGAGATCTACAGAGACATCAACACTGCTTGGCCCAAGCCCTCAGAGAGCTACCTGCTT | 911 |
| QY | 1001 | GTGGGCTATAGGACTGTGGCTGAGAGCCCTGTGTGCCAGACTTCCGAAATGTATGATTC | 1060 |
| Db | 912 | GTGGGCTATAGGACTGTGGCTGAGAGCCCTGTGTGCCAGACTTCCGAAATGTATGATTC | 971 |
| QY | 1061 | GTGGGATTTGAATTTCTGGCCCATCTTTCAGACTGGGTGTCATATCATCAACCCGAGCAG | 1120 |
| Db | 972 | GTGGGATTTGAATTTCTGGCCCATCTTTCAGACTGGGTGTCATATCATCAACCCGAGCAG | 1031 |
| QY | 1121 | GGCATGGGTACCAACGCTTTCCACACAAATGCCAGGAGTGCCTTATGTCTCCAGGCT | 1180 |
| Db | 1032 | GGCATGGGTACCAACGCTTTCCACACAAATGCCAGGAGTGCCTTATGTCTCCAGGCT | 1091 |
| QY | 1181 | TGGGGCTGGGCTTTGGCCCTATGAAAAGTTTCTCGAGCTCAGAAAAGCCGTGTGGCATGCC | 1240 |
| Db | 1092 | TGGGGCTGGGCTTTGGCCCTATGAAAAGTTTCTCGAGCTCAGAAAAGCCGTGTGGCATGCC | 1151 |
| QY | 1241 | ATGAATTTGCTGGGCTTACCGCCACGTCAGATCTGAAACAAGAACAGAAACCTGTGTGTG | 1300 |
| Db | 1152 | ATGAATTTGCTGGGCTTACCGCCACGTCAGATCTGAAACAAGAACAGAAACCTGTGTGTG | 1211 |
| QY | 1301 | GATCTTCTGTCTACCTGAGACTGTCTCTGAGCAAAATCCATAGAGGCTTGAGAGGCTTT | 1360 |

| QY | LOCUS | DEFINITION | SEQUENCE | 194 | FROM | PATENT | MO0212280. | 1271 |
|-----------------------|-----------------|--|-----------|-------------|--------|-----------------|------------|------|
| DB | LOCUS | DEFINITION | SEQUENCE | 194 | FROM | PATENT | MO0212280. | 1271 |
| QY | 1361 | GOTGCCACCTGGTGCACCCGACGACATCTCTCGATGAGTTCGAGCCCTTGCCATAT | 1420 | | | | | |
| DB | 1272 | GOTGCCACCTGGTGCACCCGACGACATCTCTCGATGAGTTCGAGCCCTTGCCATAT | 1331 | | | | | |
| QY | 1421 | CT 1422 | | | | | | |
| DB | 1332 | CT 1333 | | | | | | |
| RESULT 7 | AX381256 | Sequence 194 from Patent MO0212280. | 517 bp | DNA | Linear | PAT 18-MAR-2002 | | |
| LOCUS | AX381256 | Sequence 194 from Patent MO0212280. | 517 bp | DNA | Linear | PAT 18-MAR-2002 | | |
| DEFINITION | AX381256 | Sequence 194 from Patent MO0212280. | 517 bp | DNA | Linear | PAT 18-MAR-2002 | | |
| ACCESSION | AX381256 | Sequence 194 from Patent MO0212280. | 517 bp | DNA | Linear | PAT 18-MAR-2002 | | |
| VERSION | AX381256.1 | GI:19576075 | | | | | | |
| KEYWORDS | | | | | | | | |
| ORGANISM | human. | | | | | | | |
| SOURCE | Homo sapiens | | | | | | | |
| REFERENCE | 1 | Pyle,R.A., Xu,J. and Secrist,H. Compositions and methods for the therapy and diagnosis of colon cancer | | | | | | |
| AUTHORS | 1 | Pyle,R.A., Xu,J. and Secrist,H. Compositions and methods for the therapy and diagnosis of colon cancer | | | | | | |
| TITLE | 1 | Pyle,R.A., Xu,J. and Secrist,H. Compositions and methods for the therapy and diagnosis of colon cancer | | | | | | |
| JOURNAL | | Patent: WO 0212280-A 194 14-FEB-2002; | | | | | | |
| FEATURES | | CORIXA CORPORATION (US) | | | | | | |
| SOURCE | | Location/Qualifiers | | | | | | |
| | | 1..517 | | | | | | |
| | | /organism="Homo sapiens" | | | | | | |
| | | /db_xref="taxon:9606" | | | | | | |
| BASE COUNT | 121 a | 130 c | 124 g | 142 t | | | | |
| ORIGIN | | | | | | | | |
| Query Match | 24.9% | Score 505; | DB 6; | Length 517; | | | | |
| Best Local Similarity | 100.0% | Pred. NO. 5.6e-290; | | | | | | |
| Matches 505; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; | | | | |
| QY | 1355 | GGCTTTGCTGCACCTGCTGTCTCAGCCCTCAGCCACATCTTCTCGAATGCTTGTGAGCCCTTGC | 1414 | | | | | |
| DB | 13 | GGCTTTGCTGCACCTGCTGTCTCAGCCCTCAGCCACATCTTCTCGAATGCTTGTGAGCCCTTGC | 72 | | | | | |
| QY | 1415 | CTACATCTCTCTAGACCTTACTACATCTGTCTGTGGTATCTACACTGAGTGTGCTTCC | 1474 | | | | | |
| DB | 73 | CTACATCTCTCTAGACCTTACTACATCTGTCTGTGGTATCTACACTGAGTGTGCTTCC | 132 | | | | | |
| QY | 1475 | ACAGCTGCTCAAGCAGAGAGACTTTTGTGTCATGCTGTGTCTGAGAAACAGACTGSGG | 1534 | | | | | |
| DB | 133 | ACAGCTGCTCAAGCAGAGAGACTTTTGTGTCATGCTGTGTCTGAGAAACAGACTGSGG | 192 | | | | | |
| QY | 1535 | AACCTTATGTAGACGACACATCCACAGTAAACAGGGAATGCTCTTCTTTCTT | 1594 | | | | | |
| DB | 193 | AACCTTATGTAGACGACACATCCACAGTAAACAGGGAATGCTCTTCTTTCTT | 252 | | | | | |
| QY | 1595 | GATCTTCTGTCTGGGCGAGACTTCAGAGACCTTGTGGCTGTGAGGCCATTAAAGACAGC | 1654 | | | | | |
| DB | 253 | GATCTTCTGTCTGGGCGAGACTTCAGAGACCTTGTGGCTGTGAGGCCATTAAAGACAGC | 312 | | | | | |
| QY | 1655 | ACAGTATCAGTGAATGATGCATTAACCTCCCTGTCACATCTTGTGCCAATGGGGAATG | 1714 | | | | | |
| DB | 313 | ACAGTATCAGTGAATGATGCATTAACCTCCCTGTCACATCTTGTGCCAATGGGGAATG | 372 | | | | | |
| QY | 1715 | GATCTTACCAAGAGAGTCCACGACATTTTCCACAGAGATGCAAAATCTAGCCCTTGG | 1774 | | | | | |
| DB | 373 | GATCTTACCAAGAGAGTCCACGACATTTTCCACAGAGATGCAAAATCTAGCCCTTGG | 432 | | | | | |
| QY | 1775 | AGTTCCAGTGGATTCAGAGAGAGAGTGGGAGACAGATTGAGATGCTTATGAGCTT | 1834 | | | | | |
| DB | 433 | AGTTCCAGTGGATTCAGAGAGAGAGTGGGAGACAGATTGAGATGCTTATGAGCTT | 492 | | | | | |
| QY | 1835 | GACCATCAGACCTATCGGTAATCAG | 1859 | | | | | |

```

Db 493 GACCATCAGCTATCGTATACAG 517
|||||
RESULT 8
AX381326 389 bp DNA linear PAT 18-MAR-2002
LOCUS
DEFINITION Sequence 264 from Patent WO0212280.
ACCESSION AX381326
VERSION AX381326.1 GI:19576145
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 Pyle, R.A., Xu, J. and Secrist, H.
AUTHORS Compositions and methods for the therapy and diagnosis of colon
TITLE cancer.
JOURNAL Patent: WO 0212280-A 264 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..389
BASE COUNT 121 a 83 c 76 g 109 t
ORIGIN
Query Match 14.1%; Score 286; DB 6; Length 389;
Best Local Similarity 100.0%; Pred. No. 9, 2e-159;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1732 CTCACACGATTTTCCACAGATGCAATTTGAGCCCTTGAGTCCAGTGGATCA 1791
|||||
Db 13 CTCACACGATTTTCCACAGATGCAATTTGAGCCCTTGAGTCCAGTGGATCA 72
|||||
Qy 1792 AGGAGAGAGTGGGAGACAGGTGGATGCTTATGAGTATGACATACAGCTATCG 1851
|||||
Db 73 AGGAGAGAGTGGGAGACAGGTGGATGCTTATGAGTATGACATACAGCTATCG 132
|||||
Qy 1852 GTAATCAGAAATATGAAACAAATCTCTGCACAAAGAGAGAGCTTTAAGTTCAGGG 1911
|||||
Db 133 GTAATCAGAAATATGAAACAAATCTCTGCACAAAGAGAGAGCTTTAAGTTCAGGG 192
|||||
Qy 1912 TGCCGCGGCTGATTTGATATCATCTCCCTGCAATTTCCGATACATAGAGACTT 1971
|||||
Db 193 TGCCGCGGCTGATTTGATATCATCTCCCTGCAATTTCCGATACATAGAGACTT 252
|||||
Qy 1972 TGACCTGTGAAGCTGCATCTGTTAATCTAAATTTCCCAATTAAG 2017
|||||
Db 253 TGACCTGTGAAGCTGCATCTGTTAATCTAAATTTCCCAATTAAG 298
|||||

RESULT 9
AF176839 1462 bp DNA linear PRI 22-SEP-1999
LOCUS
DEFINITION Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase
ACCESSION AF176839
VERSION AF176839.1 GI:5917707
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1462)
AUTHORS Lee, J.K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
TITLE Cloning and characterization of a mammalian
N-acetylglucosamine-6-sulfotransferase that is highly restricted to
intestinal tissue
JOURNAL Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
MEDLINE 99423499
PUBMED 10491328

```

```

REFERENCE 2 (bases 1 to 1462)
AUTHORS Lee, J.K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA
FEATURES
source Location/Qualifiers
1..1462
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1"
/tissue_type="intestine"
<229..>1401
/gene="I-GlcNAc-6-ST"
<229..>1401
/gene="I-GlcNAc-6-ST"
/product="N-acetylglucosamine 6-O-sulfotransferase"
229..1401
/gene="I-GlcNAc-6-ST"
/feature="possibly involved in biosynthesis of mucosal
ligands for L-selectin"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD56001.1"
/db_xref="GI:5917708"
/translation="WMLEPSSKTYVILLAQTCILFLISRRGSSPPAGDEPRVH
LVLSWRSGSSFLGQLFSQHPDVEFLIEPAMWVITLSQGSATLHMAVRLNRSITL
CDMDVEFVAMPQSRNLSAFENWATSRALCPSPACAPFPGTISQDYCKILCTROPIS
LAREACRSYSHVVLKEVFEFLNQLVPLDSPALNRLVHLVPRVRLVLSREAGPI
LARDGIVLGTNGKVEADPHRLRLREVCGRSHVLAELATLPPFRLGRVRLVFEFD
LAREPLAIRALVAFETGLTLPOLAEAHNTTHSGIGKPEATKESRNARNVSCWM
RHAPETKILVOPKACALDILCYRPPVSYSDQQRDLTLDLVLRGDRHSWASPD"

BASE COUNT 239 a 510 c 446 g 267 t
ORIGIN
Query Match 2.9%; Score 59; DB 9; Length 1462;
Best Local Similarity 100.0%; Pred. No. 1e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 678 AGGCTGCGCCCTCTACACGAGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 736
|||||
Db 719 AGGCTGCGCCCTCTACACGAGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 777
|||||

RESULT 10
AX327330 1647 bp DNA linear PAT 07-JAN-2002
LOCUS
DEFINITION Sequence 15 from Patent WO0179468.
ACCESSION AX327330
VERSION AX327330.1 GI:18097876
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 Pollock, J.L., Hafalia, A., Burford, N., Ring, H.2., Lai, P.,
Tribouley, C.M., Yao, M.G., Yue, H., Tang, Y.T., Patterson, C., Das, D.,
Sanjanwala, M.S., Gandhi, A.R., Reddy, R., Khan, F.A., Baughn, M.R.,
Ranummar, J., Griffin, J.A. and Au-Yang, J.
TITLE Drug metabolizing enzymes
JOURNAL Patent: WO 0179468-A 15 25-OCT-2001;
Incyte Genomics, Inc. (US)
FEATURES
source Location/Qualifiers
1..1647
/organism="Homo sapiens"
/db_xref="taxon:9606"
/feature="Incyte ID NO: 7472777CBI"

BASE COUNT 287 a 560 c 501 g 299 t
ORIGIN
Query Match 2.9%; Score 59; DB 6; Length 1647;
Best Local Similarity 100.0%; Pred. No. 1e-22;

```

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 678 AGGCTGCCGCTCTACAGCAGCGTGTCTCAAGAGGTGCGCTTCTTCAACCTGCAG 736
|||||
Db 488 AGGCTGCCGCTCTACAGCAGCGTGTCTCAAGAGGTGCGCTTCTTCAACCTGCAG 546

RESULT 11
AF176838 2170 bp mRNA linear PRI 22-SEP-1999
LOCUS Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase
DEFINITION (1-GlcNAc-6-ST) mRNA, complete cds.
ACCESSION AF176838
VERSION AF176838.1 GI:5917705
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2170)
AUTHORS Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
TITLE Cloning and characterization of a mammalian
N-acetylglucosamine-6-sulfotransferase that is highly restricted to
intestinal tissue
JOURNAL Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
MEDLINE 99423499
PubMed 10491328
REFERENCE 2 (bases 1 to 2170)
AUTHORS Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA

FEATURES
Source Location/Qualifiers
1..2170
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1"
/tissue_type="intestine"
1..2170
/gene="1-GlcNAc-6-ST"
344..1516
/gene="1-GlcNAc-6-ST"
/note="possibly involved in biosynthesis of mucosal
ligands for L-selection"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAB5600.1"
/db_xref="GI:5917706"
/translation="MWLPRESSKTVTLVLLAOTCLLEFIISPGSSPAGGEDRVHV
LVLSWRSGLSFLGOLFQSDVFLMEPMHWITLISQSAATLHMAVADLMRSIFL
CDMDVPAIYPMRNLSDLEFQWAVSRALCSPPACSAFPRRTISKDVCIKICTROFS
LARDACRSYSHVVLKEVRFNLQVILPILSDPLNRIYHVLVDPRAVLRSGRAAGPI
LARDNGVILGTNGWVADPHRLIRVGVCSHRIAEALIKRPPILRGYRVREDD
LAREPLAEIRALYAFITGLITTPLEAMNHNITGSGGRIETEFHHSRRARVNSGAM
RHLPFKIRRVDELCAGLQLGIRPVISADQRLTLDVLPRGPDHFSWASPD"

BASE COUNT 449 a 682 c 612 g 427 t

Query Match 2.9%; Score 59; DB 9; Length 2170;
Best Local Similarity 100.0%; Pred. No. 1e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 678 AGGCTGCCGCTCTACAGCAGCGTGTCTCAAGAGGTGCGCTTCTTCAACCTGCAG 736
|||||
Db 834 AGGCTGCCGCTCTACAGCAGCGTGTCTCAAGAGGTGCGCTTCTTCAACCTGCAG 892

RESULT 12
AF219990 2544 bp mRNA linear PRI 26-OCT-2000
LOCUS Homo sapiens corneal N-acetylglucosamine-6-O-sulfotransferase
DEFINITION (CHST6) mRNA, complete cds.
REFERENCE (CHST6) mRNA, complete cds.

ACCESSION AF219990
VERSION AF219990.1 GI:11023145
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2544)
AUTHORS Akama,T.O., Nishida,K., Nakayama,A., Watanabe,H., Fujiwara,T.,
Nakanura,T., Dots,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,
Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
TITLE Macular corneal dystrophy type I and type II are caused by distinct
mutations in a new sulfotransferase gene
JOURNAL Nat. Genet. 26 (2), 237-241 (2000)
MEDLINE 20472330
PubMed 11017086
REFERENCE 2 (bases 1 to 2544)
AUTHORS Akama,T.O. and Fukuda,M.N.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-1999) Glycobiology Program, The Burnham
Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
Source Location/Qualifiers
1..2544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22"
1..2544
/gene="CHST6"
/note="responsible for macular corneal dystrophy"
693..1880
/gene="CHST6"
/function="transfers sulfate to 6-OH of GlcNAc in keratan
sulfate"
/note="C-GlcNAc6ST"
/codon_start=1
/product="corneal
N-acetylglucosamine-6-O-sulfotransferase"
/protein_id="AAG26325.1"
/db_xref="GI:11023146"
/translation="MWLPVSVSTAVTALLAOTFLLELVSPGSSPAGGEARVHL
VLWSRSGSSEFGOLFNOHPDVFLMEPMHWITLISQSAATLHMAVADLMRSIFL
DMDVPAIYPMRNLSDLEFQWAVSRALCSPPACSAFPRRTISKDVCIKICTROFS
LARDACRSYSHVVLKEVRFNLQVILPILSDPLNRIYHVLVDPRAVLRSGRAAGPI
LARDNGVILGTNGWVADPHRLIRVGVCSHRIAEALIKRPPILRGYRVREDD
LAREPLAEIRALYAFITGLITTPLEAMNHNITGSGGRIETEFHHSRRARVNSGAM
RHLPFKIRRVDELCAGLQLGIRPVISADQRLTLDVLPRGPDHFSWASPD"

BASE COUNT 460 a 799 c 733 g 552 t

Query Match 2.9%; Score 59; DB 9; Length 2544;
Best Local Similarity 100.0%; Pred. No. 1e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 678 AGGCTGCCGCTCTACAGCAGCGTGTCTCAAGAGGTGCGCTTCTTCAACCTGCAG 736
|||||
Db 1180 AGGCTGCCGCTCTACAGCAGCGTGTCTCAAGAGGTGCGCTTCTTCAACCTGCAG 1238

RESULT 13
AF246718 3278 bp mRNA linear PRI 31-OCT-2000
LOCUS Homo sapiens intestinal GlcNAc-6-sulfotransferase (CHST5) mRNA,
DEFINITION complete cds, alternatively spliced.
ACCESSION AF246718
VERSION AF246718.1 GI:11055254
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3278)

| REFERENCE | TITLE | FEATURES |
|---|---|----------|
| JOURNAL MEDLINE PUBMED 21096027 1181564 | Chromosomal localization and genomic organization for the galactose/ N-acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene family Glycobiology 11 (1), 75-87 (2001) | |
| AUTHORS ROSEN,S.D. | 2 (bases 1 to 3786) | |
| HEMMERICH,S., BHAKTA,S., LEE,J.-K., BISTRUP,A., RIDDLE,N.R. and JOORNAL | Direct Submission Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA | |
| FEATURES | Location/Qualifiers | |
| SOURCE | 1..3786 | |
| gene | /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /map="16q23.1" 1..3786 | |
| CDS | /gene="GST4beta" /note="clusters with the intestinal N-acetylglucosamine 6-O-sulfotransferase (GST4beta) gene in GenBank Accession Numbers AF176838 and AF176839" 189..1376 /gene="GST4beta" /note="similar to N-acetylglucosamine 6-O-sulfotransferase GST-4alpha" /codon_start=1 /product="N-acetylglucosamine 6-O-sulfotransferase GST-4beta" /protein_id="AA648244.1" /db_xref="GI:12060804" /translation="MMLPRVSTAVTALLAQTFLLFLVSRPPSSPAGEARVHVLVSRRGSSFFVGLEFQHDVEFLMEPAHWTTISGSAITLHMAFDLVASFICDMVDVADILPMRNSDLPQWASRALCSPACAPRGASISSEAVCKPACARQFPLARECRYSHVVLKEVRFNTLOVILPLSDPALNIRVHVRPRAVLREROTAKALARDGILVLGNGTINVEADPEGLRVYREYCRSHVYIADEATIKPPLRLGRIRLYPEDLAREDELAIRALVATFAGTSLTPQLEAMVHNTTHSGPEAREAKRTSRNLNVSQARHAPFAFKIRRVQELCAGALQILGLRVYSDEQRNLALDLVPLRGLNGFTWASSTASH PRN" | |
| polyA_signal | 3106..3111 /gene="GST4beta" /note="alternative signal" | |
| polyA_signal | 3734..3739 /gene="GST4beta" | |
| BASE COUNT | 848 a 1030 c 1104 g 804 t | |
| ORIGIN | | |
| Query Match | 2.9%: Score 59; DB 9; Length 3786; | |
| Best Local Similarity | 100.0%; Pred. No. 1,1e+22; | |
| Matches | 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | |
| QY 678 | AGGCGTGCCTCTCTACAGCCAGCGTGTCCTCAAGAGAGGTGCGCTTCTCAACCTGACAG 736 | |
| b 676 | AGGCGTGCCTCTCTACAGCCAGCGTGTCCTCAAGAGAGGTGCGCTTCTCAACCTGACAG 734 | |
| RESULT 15 | | |
| LOCUS | AF219991 | |
| DEFINITION | Homo sapiens intestinal N-acetylglucosamine-6-O-sulfotransferase (CHST5) and corneal N-acetylglucosamine-6-O-sulfotransferase (CHST6) genes, complete cds. | |
| ACCESSION | AF219991 | |
| VERSION | AF219991.1 | |
| KEYWORDS | GI:11023147 | |
| SOURCE | | |
| ORGANISM | Homo sapiens. | |
| REFERENCE | Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Euthera: Primates: Catarrhini: Hominoidea: Homo. 1 (bases 1 to 71503) | |
| AUTHORS | Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., | |

TITLE Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
 Macular corneal dystrophy type I and type II are caused by distinct
 mutations in a new sulphotransferase gene
 JOURNAL Nat. Genet. 26 (2), 237-241 (2000)
 MEDLINE 20472330
 PUBMED 11017086
 REFERENCE 2 (bases 1 to 71503)
 AUTHORS AKama,T.O. and Fukuda,M.N.
 TITLE Direct Submission
 JOURNAL Submitted (29-DEC-1999) Glycobiology Program, The Burnham
 Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES Location/Qualifiers
 source 1..71503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q22"
 /clone="CITB-483K2"
 358..461
 /rpt_family="L2"
 /rpt_type=dispersed
 complement(462..771)
 /rpt_family="Alu"
 /rpt_type=dispersed
 772..945
 /rpt_family="L2"
 /rpt_type=dispersed
 1169..1338
 /rpt_family="MIR"
 /rpt_type=dispersed
 complement(1353..1804)
 /rpt_family="Alu"
 /rpt_type=dispersed
 2616..2916
 /rpt_family="Alu"
 /rpt_type=dispersed
 3309..3532
 /note="similar to EST H12297 (EST cluster Hs.31147)"
 3865..4028
 /note="similar to EST H12297 (EST cluster Hs.31147)"
 4039..4406
 /rpt_family="Alu"
 /rpt_type=dispersed
 4793..5250
 /rpt_family="L2"
 /rpt_type=dispersed
 complement(5251..5329)
 /rpt_family="MIR"
 /rpt_type=dispersed
 complement(5768..6080)
 /rpt_family="Alu"
 /rpt_type=dispersed
 complement(6081..6265)
 /rpt_family="Alu"
 /rpt_type=dispersed
 6111..6567
 /note="similar to EST AA001593 (EST cluster Hs.6853)"
 8035..8350
 /rpt_family="Alu"
 /rpt_type=dispersed
 10545..11049
 /note="similar to EST A1950923 (EST cluster Hs.6853)"
 11043..11195
 /rpt_family="Alu"
 /rpt_type=dispersed
 11200..11222
 /rpt_family="Simple-repeat"
 11570..11871
 /rpt_family="Alu"
 /rpt_type=dispersed
 complement(12065..12198)
 /rpt_family="L1"

repeat_region complement(12205..12493)
 /rpt_family="Alu"
 /rpt_type=dispersed
 complement(12813..13110)
 /rpt_family="Alu"
 /rpt_type=dispersed
 14185..20821
 /gene="CHST5"
 mRNA join(14185..14228,16169..16263,17712..20821)
 /gene="CHST5"
 /product="intestinal N-acetylglucosamine
 6-O-sulfotransferase"
 14681..14713
 /rpt_family="Simple-repeat"
 /rpt_type=tandem
 complement(14715..14990)
 /rpt_family="Alu"
 /rpt_type=dispersed
 complement(14995..15288)
 /rpt_family="Alu"
 /rpt_type=dispersed
 complement(15292..15589)
 /rpt_family="Alu"
 /rpt_type=dispersed
 15668..15718
 /rpt_family="Alu"
 /rpt_type=dispersed
 15977..16098
 /rpt_family="MIR"
 /rpt_type=dispersed
 complement(17298..17433)
 /rpt_family="MIR"
 /rpt_type=dispersed
 complement(18001..18310)
 /rpt_family="Alu"
 /rpt_type=dispersed
 complement(18420..18591)
 /rpt_family="MIR"
 /rpt_type=dispersed
 18732..18767
 /rpt_family="Simple-repeat"
 /rpt_type=tandem
 18967..20202
 /gene="CHST5"
 /note="I-GlcNAc6ST; carbohydrate sulfoyltransferase 5;
 expressed in the small intestine and colon"
 /codon_start=1
 /product="intestinal
 N-acetylglucosamine-6-O-sulfotransferase"
 /protein_id="AA626326.1"
 /db_xref="GI:11023148"
 /translation="MGMRARVPKVAHSTRPPAPARMWLPRESSKTIVTILIAQTCLL
 LFILSRCPSPAGCHDRVHVLVLSMRSGSFLQLFSQHDVYLMEPAHWVTL
 SOGSATILMAVRDLMSIFLDMVFPDAVPMOSNLISAFEMATSAICSPACSAF
 PRGTSIKDVCCTICITROPEFLAREACREYSHVVLKEVEFFLVLYLSPRLNTR
 IYHLVRPAVLRSRFAGPITARDNGIYLGNGKWEKDPRLRIREYCRSHVIAE
 AATLKPPLRLGRYRLVREDFLAREPLAIRLVLTGTLTPOLEAMTHNTGSGI
 GKPIEAFHSSRNARVSOAMRHALPFTKILRVQVCAGALLGYPVYSADQGRDL
 TLDIVLRGPDHFSNASPD"
 20504..20806
 /rpt_family="Alu"
 /rpt_type=dispersed
 20863..20961
 /rpt_family="L1"
 /rpt_type=dispersed
 20962..21200
 /rpt_family="MER4-group"
 /rpt_type=dispersed
 21325..21429
 /rpt_family="MER4-group"
 /rpt_type=dispersed
 21430..21722
 repeat_region

```

/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 21725..22207
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 22008..22040
/rpt_family="Simple_repeat"
/rpt_type=tandem
repeat_region 22041..22313
/rpt_family="MER3-group"
/rpt_type=dispersed
repeat_region 22325..22593
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 22594..22949
/rpt_family="MER4-group"
/rpt_type=dispersed
repeat_region 23116..23221
/rpt_family="Simple_repeat"
/rpt_type=tandem
repeat_region 23344..23400
/rpt_family="MaLR"
/rpt_type=dispersed
repeat_region 23433..23456
/rpt_family="Simple_repeat"
/rpt_type=tandem
repeat_region complement(23457..23720)
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region complement(23789..24140)
/rpt_family="MER21-group"
/rpt_type=dispersed
repeat_region 24141..24287
/rpt_family="MaLR"
/rpt_type=dispersed
repeat_region complement(24308..24420)
/rpt_family="MaLR"
/rpt_type=dispersed
repeat_region complement(24467..24528)
/rpt_family="MER1_type"

Query Match 2.9%; Score 59; DB 9; Length 71503;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCGTCGGCTCTACAGCAGCGTGTGCTCAAGAGGTCGCTTCAACCTGCAG 736
|||||
Db 19520 AGGCGTCGGCTCTACAGCAGCGTGTGCTCAAGAGGTCGCTTCAACCTGCAG 19578

RESULT 16
AC009163/ 157337 bp DNA linear PRI 02-NOV-2001
LOCUS Homo sapiens chromosome 16 clone RP11-77K12, complete sequence.
DEFINITION AC009163
AC009163.5 GI:16596526
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 157337)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157337)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 157337)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission

```

```

JOURNAL Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 2, 2001 this sequence version replaced gi:7689976.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
Location/Qualifiers
1..157337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-77K12"
BASE COUNT 45295 a 36632 c 36308 g 39102 t
ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 157337;
Best Local Similarity 100.0%; Pred. No. 1.3e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCGTCGGCTCTACAGCAGCGTGTGCTCAAGAGGTCGCTTCAACCTGCAG 736
|||||
Db 144095 AGGCGTCGGCTCTACAGCAGCGTGTGCTCAAGAGGTCGCTTCAACCTGCAG 144037

RESULT 17
AC025287 194832 bp DNA linear PRI 26-JAN-2002
LOCUS Homo sapiens chromosome 16 clone RP11-490B18, complete sequence.
DEFINITION AC025287
AC025287.8 GI:18376863
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jan 26, 2002 this sequence version replaced gi:17976465.
Draft Sequence Produced by DOE Joint Genome Institute

```



```
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.4% of Sequence;
Estimated Total Number of Errors is 0.6.

FEATURES
  source
    1. 194832
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /chromosome="16"
       /clone="RP11-490B18"

BASE COUNT   53784 a 43240 c 43815 g 53993 t
ORIGIN
Query Match      2.9%; Score 59; DB 9; Length 194832;
Best Local Similarity 100.0%; Pred. No. 1.3e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY   678 AGGCGTCCGCTCTCAAGCAGCTGTGCTCAAGAGGCGCTTCTTCAACCTGCAG 736
Db 188038 AGGCGTCCGCTCTCAAGCAGCTGTGCTCAAGAGGCGCTTCTTCAACCTGCAG 188096

RESULT 18
AC009105      208185 bp   DNA   linear   HTG 25-APR-2001
LOCUS
DEFINITION   Homo sapiens chromosome 16 clone RP11-455E15, WORKING DRAFT
ACCESSION   AC009105
VERSION     AC009105.7   GI:13786304
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 208185)
            DOE Joint Genome Institute.
            Sequencing of Human Chromosome 16
            Unpublished
            2 (bases 1 to 208185)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Apr 25, 2001 this sequence version replaced gi:1689928.
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RP11-455E15
-----
Summary Statistics
Consensus quality: 180040 bases at least Q40
Consensus quality: 192095 bases at least Q30
Consensus quality: 195726 bases at least Q20
Estimated insert size: 195650; agarose-ef estimation
Estimated coverage: 6.57 in Q20 bases; agarose-ef estimation
Quality coverage: 6.23 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
      consists of 17 contigs. The true order of the pieces
      is not known and their order in this sequence record is
      arbitrary. Gaps between the contigs are represented as
      runs of N, but the exact sizes of the gaps are unknown.
      This record will be updated with the finished sequence
      as soon as it is available and the accession number will
      be preserved.
      1 1062: contig of 1062 bp in length
      * 1063 1162: gap of unknown length
      * 1163 2359: contig of 1197 bp in length

www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.4% of Sequence;
Estimated Total Number of Errors is 0.6.

FEATURES
  source
    1. 194832
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /chromosome="16"
       /clone="RP11-490B18"

BASE COUNT   53784 a 43240 c 43815 g 53993 t
ORIGIN
Query Match      2.9%; Score 59; DB 2; Length 208185;
Best Local Similarity 100.0%; Pred. No. 1.3e-22;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY   678 AGGCGTCCGCTCTCAAGCAGCTGTGCTCAAGAGGCGCTTCTTCAACCTGCAG 736
Db 107982 AGGCGTCCGCTCTCAAGCAGCTGTGCTCAAGAGGCGCTTCTTCAACCTGCAG 108040

RESULT 19
AR203340      37 bp   DNA   linear   PAT 20-JUN-2002
LOCUS
DEFINITION   Sequence 7 from patent US 6365365.
ACCESSION   AR203340
VERSION     AR203340.1   GI:21499704
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 37)
            Bistrup, A., Rosen, S.D., Tangemann, K. and Hemmerich, S.
            Method of determining whether an agent modulates glycosyl
            sulfotransferase-3
            Patent: US 6365365-A 7 02-APR-2002;
            Location/Qualifiers
            source
              1. 37
                 /organism="unknown"

BASE COUNT   14 a 9 c 8 g 6 t
ORIGIN
Query Match      1.8%; Score 37; DB 6; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 884 AACTCAAGAGAGACCAACCTACTATGTATGC 920
 AF109155
 DB 1 AACTCAAGAGAGACCAACCTACTATGTATGC 37

RESULT 20
 AF131236 1926 bp DNA linear ROD 10-AUG-1999
 LOCUS AF131236
 DEFINITION Mus musculus N-acetylglucosamine 6-O-sulfotransferase gene,
 complete cds.

ACCESSION AF131236
 VERSION AF131236.1 GI:4927115

KEYWORDS Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1926)
 Bistup,A., Bhakta,S., Lee,J.K., Belov,Y.Y., Gunn,M.D., Zuo,F.R.,
 Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerlich,S.
 Sulfotransferases of two specificities function in the
 reconstitution of high endothelial cell ligands for L-selectin
 J. Cell Biol. 145 (4), 899-910 (1999)

TITLE JOURNAL
 MEDLINE 99264336
 PUBMED 10330415

REFERENCE 2 (bases 1 to 1926)
 Bistup,A., Tangemann,K., Bhakta,S., Lee,J.K., Belov,Y.Y.,
 Gunn,M.D., Zuo,F.R., Huang,C.C., Kannagi,R., Rosen,S.D. and
 Hemmerlich,S.
 Direct Submission
 Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Avenue, Palo Alto, CA 94304, USA

FEATURES
 Source
 1..1926
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /cell_type="high endothelial"
 /tissue_type="tissue"
 <405..>1571
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 405..1571
 /function="Biosynthesis of endothelial ligands for
 L-selectin"
 /note="HEC-GLCNAc-6-Sr"
 /codon_start=1
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 /protein_id="AAD33016.1"
 /db_xref="GI:4927116"
 /translation="MMLIKGRLLMFLGSOVIVVAFIMSVHRHLSQRESRPVAH
 LVLSMRSGSEFVGOLFQHPDVEYLMPEAMHWMTTSTAMKILHMAVRDLRSVFL
 CMSSVFDAYMNPGRKOSLFWMEOSRALCPVCDFFRAHFISSPKKCKLIGQGF
 DMVEKACRSHGFVLEKEVRFSLQALPLIDPSLNHVLVRDPRVAFRSREHTTI
 ELMBDSHVLVGHLETLKEEDQPYAMKTIKSOYDIKALQTLFEALQQRFLRYE
 ELVAPLAQTRLRYKFGVLDLPLQATWYVNRGKMGQAHFNNARNANVSGAMR
 WSLPEYKVSQLDACGFAMDGLGVQVSQCGNLSIDLSSSHILQVFRFG"

BASE COUNT 426 a 520 c 484 g 496 t

ORIGIN

Query Match 1.3%; Score 27; DB 10; Length 1926;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 GACATGAGCGCTTTGATGCTACATG 520
 AF109155
 DB 714 GACATGAGCGCTTTGATGCTACATG 740

RESULT 21
 AF109155 2201 bp mRNA linear ROD 27-JUL-1999
 LOCUS AF109155
 DEFINITION Mus musculus L-selectin ligand sulfotransferase mRNA, complete cds.

ACCESSION AF109155
 VERSION AF109155.1 GI:5596405
 KEYWORDS Mus musculus.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2201)
 Hiraoka,N., Petryniak,B., Nakayama,J., Tsuboi,S., Suzuki,M.,
 Yen,J.C., Izawa,D., Tanaka,T., Miyasaka,M., Lowe,J.B. and Fukuda,M.
 A novel, high endothelial venule-specific sulfotransferase
 expresses 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed
 by CD34
 Immunity 11 (1), 79-89 (1999)

TITLE JOURNAL
 MEDLINE 99361934
 PUBMED 10435581

REFERENCE 2 (bases 1 to 2201)
 Hiraoka,N. and Fukuda,M.
 Direct Submission
 Submitted (23-NOV-1998) Glycobiology, The Burnham Institute, 10901
 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
 source
 1..2201
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 319..1485
 /function="directs expression of 6-sulfo sialyl Lewis X,
 an L-selectin ligand displayed by CD34"
 /note="155r"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAD45579.1"
 /db_xref="GI:5596406"
 /translation="MMLIKGRLLMFLGSOVIVVAFIMSVHRHLSQRESRPVAH
 LVLSMRSGSEFVGOLFQHPDVEYLMPEAMHWMTTSTAMKILHMAVRDLRSVFL
 CMSSVFDAYMNPGRKOSLFWMEOSRALCPVCDFFRAHFISSPKKCKLIGQGF
 DMVEKACRSHGFVLEKEVRFSLQALPLIDPSLNHVLVRDPRVAFRSREHTTI
 ELVAPLAQTRLRYKFGVLDLPLQATWYVNRGKMGQAHFNNARNANVSGAMR
 WSLPEYKVSQLDACGFAMDGLGVQVSQCGNLSIDLSSSHILQVFRFG"

BASE COUNT 537 a 629 c 534 g 501 t

ORIGIN

Query Match 1.3%; Score 27; DB 10; Length 2201;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 GACATGAGCGCTTTGATGCTACATG 520
 AF109155
 DB 628 GACATGAGCGCTTTGATGCTACATG 654

RESULT 22
 AF302109 1398 bp mRNA linear PRI 20-SEP-2000
 LOCUS AF302109
 DEFINITION Homo sapiens ps20 WAF-type four-disulfide core domain protein mRNA,
 complete cds.

ACCESSION AF302109
 VERSION AF302109.1 GI:10198226

KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1398)
 Ong,C.K., Ng,C.Y., Lim,K.B., Chan,T.W.M.G. and Huynh,H.
 Molecular Cloning and Characterization of the Human PS20 protein in
 Human Uterus
 Unpublished

TITLE JOURNAL
 MEDLINE 2 (bases 1 to 1398)
 PUBMED 10435581

REFERENCE 2 (bases 1 to 1398)
 Ong,C.K., Ng,C.Y., Lim,K.B., Chan,T.W.M.G. and Huynh,H.
 Direct Submission
 Submitted (01-SEP-2000) Cellular And Molecular Research, National


```

* 32255 33204: contig of 950 bp in length
*      gap of unknown length
* 33205 34156: contig of 952 bp in length
*      gap of unknown length
* 34157 35051: contig of 895 bp in length
*      gap of unknown length
* 35052 36006: contig of 955 bp in length
*      gap of unknown length
* 36007 36955: contig of 949 bp in length
*      gap of unknown length
* 36956 37899: contig of 944 bp in length
*      gap of unknown length
* 37900 38880: contig of 981 bp in length
*      gap of unknown length
* 38881 39843: contig of 963 bp in length
*      gap of unknown length
* 39844 40789: contig of 946 bp in length
*      gap of unknown length
* 40790 41696: contig of 907 bp in length
*      gap of unknown length
* 41697 42487: contig of 791 bp in length
*      gap of unknown length
* 42488 43448: contig of 961 bp in length
*      gap of unknown length
* 43449 44434: contig of 986 bp in length
*      gap of unknown length
* 44435 45348: contig of 914 bp in length
*      gap of unknown length
* 45349 46306: contig of 958 bp in length
*      gap of unknown length
* 46307 47248: contig of 942 bp in length
*      gap of unknown length
* 47249 48181: contig of 933 bp in length
*      gap of unknown length
* 48182 49144: contig of 963 bp in length
*      gap of unknown length
* 49145 50101: contig of 957 bp in length
*      gap of unknown length
* 50102 51049: contig of 948 bp in length
*      gap of unknown length
* 51050 51982: contig of 933 bp in length
*      gap of unknown length
* 51983 52928: contig of 946 bp in length
*      gap of unknown length
* 52929 53867: contig of 939 bp in length
*      gap of unknown length
* 53868 54843: contig of 976 bp in length
*      gap of unknown length
* 54844 55755: contig of 912 bp in length
*      gap of unknown length
* 55756 56703: contig of 948 bp in length
*      gap of unknown length
* 56704 57660: contig of 957 bp in length
*      gap of unknown length
* 57661 58593: contig of 933 bp in length
*      gap of unknown length
* 58594 59547: contig of 954 bp in length
*      gap of unknown length
* 59548 60496: contig of 949 bp in length
*      gap of unknown length
* 60497 61450: contig of 954 bp in length
*      gap of unknown length
* 61451 62409: contig of 959 bp in length
*      gap of unknown length
* 62410 63368: contig of 959 bp in length
*      gap of unknown length
* 63369 64320: contig of 952 bp in length
*      gap of unknown length
* 64321 65317: contig of 997 bp in length
*      gap of unknown length
* 65318 66295: contig of 978 bp in length
*      gap of unknown length
* 66296 67201: contig of 906 bp in length

```

```

Query Match      1.2%; Score 24; DB 2; Length 79598;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2008 CCCAATTAAGAAAAA 2031
Db 60724 CCCAATTAAGAAAAA 60701

RESULT 24
AC095651/C
LOCUS
DEFINITION
ACCESSION
AC095651.3 GI:21722647
VERSION
HTG: HTGS PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 99395)
Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
Alstrooks S.L., Amaralunge H.C., Are J.R., Ayale M., Banks T.,
Barbaita V., Benton J., Bimaye K., Blankensbury K., Bonnin D.,
Bouck J., Bowle S., Brileva M., Brown E., Brown M., Bryant N.P.,
Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,
Carton T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhury I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Devila M.L., Davis C., Davy-Carroll L., Dederich D.A.,
DeLaney K.R., Delgado O., Denn A.L., Ding T., Dinh H.H.,
Douthwaite K.J., Dreper H., Dugan-Rocha S., Durbin K.J.,
Earhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorelli J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogue M., Holloway C., Hollins B.,
Homsli F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karlsone E., Kelly S., Khan U., King L., Korvah J., Kovar C.,
Kratovic J., Kurishi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Louisseged H.,
Lozada R.T., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maneshwari M., Mapa P., Martin R., Martindale A., Martinez E.,
Massey E., Mawhinley E., McLeod M.P., Meador M., Nel G., Metzker M.,
Miner G., Miner Z., Mitchell T., Monabati K., Morgan M., Morris S.,
Mosier M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
Nguyen N., Nickerson E., Nwokkenko S., Oguni M., Okunolu G.,
Oragunye N., Oyiedo R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
Rivers M., Rojas A., Rojupokan I., Rolfe M., Ruiz S., Savery G.,
Scherrer S., Scott G., Shen H., Shoostitani N., Sisson I.,
Sodergren E., Sotaike T., Sparks A., Stanley H., Stone H.,
Sutton A., Syatek A., Tabor P., Tameis A., Tameis K., Tang H.,
Tansey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S.,
Usmani K., Vasquez L., Vera V., Villalón D., Vinson R., Wang Q.,

```

Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 99395)
 Worley, K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 99395)
 Worley, K.C.
 Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced g1:17942210.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc@bcm.tmc.edu
 ----- Project Information
 Center project name: GCXS
 Center clone name: CH230-8E2
 ----- Summary Statistics
 Sequencing vector: plasmid
 Chemistry: Dye-terminator Big Dye 100% of reads
 Assembly program: Phrap, version 0.990329
 Consensus quality: 66825 bases at least Q40
 Consensus quality: 71670 bases at least Q30
 Consensus quality: 73911 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 44 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1425: contig of 1425 bp in length
 * 1426 1525: gap of unknown length
 * 1526 2919: contig of 1394 bp in length
 * 2920 3019: gap of unknown length
 * 3020 4525: contig of 1506 bp in length
 * 4526 4625: gap of unknown length
 * 4626 5831: contig of 1206 bp in length
 * 5832 5931: gap of unknown length
 * 5932 6935: contig of 1004 bp in length
 * 6936 7035: gap of unknown length
 * 7036 8283: contig of 1248 bp in length
 * 8284 8383: gap of unknown length
 * 8384 9805: contig of 1422 bp in length
 * 9806 9905: gap of unknown length
 * 9906 11125: contig of 1221 bp in length
 * 11127 11226: gap of unknown length
 * 11227 12713: contig of 1487 bp in length
 * 12714 12813: gap of unknown length
 * 12814 13930: contig of 1117 bp in length
 * 13931 14030: gap of unknown length
 * 14031 15456: contig of 1426 bp in length
 * 15457 15556: gap of unknown length
 * 15567 16890: contig of 1334 bp in length
 * 16891 16990: gap of unknown length
 * 16991 18008: contig of 1018 bp in length
 * 18009 18108: gap of unknown length
 * 18109 19518: contig of 1410 bp in length
 * 19519 19618: gap of unknown length

19619 20891: contig of 1273 bp in length
 20892 20991: gap of unknown length
 20992 23151: contig of 2160 bp in length
 23152 23251: gap of unknown length
 23252 24301: contig of 1049 bp in length
 24301 24400: gap of unknown length
 24401 25937: contig of 1537 bp in length
 25938 26037: gap of unknown length
 26038 27412: contig of 1375 bp in length
 27413 27512: gap of unknown length
 27513 28754: contig of 1222 bp in length
 28735 28834: gap of unknown length
 28835 29956: contig of 1122 bp in length
 29957 30056: gap of unknown length
 30057 31347: contig of 1291 bp in length
 31348 31447: gap of unknown length
 31448 32624: contig of 1177 bp in length
 32625 32724: gap of unknown length
 32725 34035: contig of 1311 bp in length
 34036 34135: gap of unknown length
 34136 35432: contig of 1297 bp in length
 35433 35532: gap of unknown length
 35533 37135: contig of 1603 bp in length
 37136 37235: gap of unknown length
 37236 39141: contig of 1906 bp in length
 39142 40831: gap of unknown length
 40832 40931: contig of 1590 bp in length
 40932 43324: contig of 2393 bp in length
 43325 43424: gap of unknown length
 43425 45280: contig of 1866 bp in length
 45281 45390: gap of unknown length
 45391 47689: contig of 2279 bp in length
 47690 47769: gap of unknown length
 47770 49684: contig of 1895 bp in length
 49685 49764: gap of unknown length
 49765 52591: contig of 2827 bp in length
 52592 52691: gap of unknown length
 52692 55345: contig of 2654 bp in length
 55346 55445: gap of unknown length
 55446 58258: contig of 2813 bp in length
 58259 58358: gap of unknown length
 58359 60686: contig of 2328 bp in length
 60687 60786: gap of unknown length
 60787 63505: contig of 2719 bp in length
 63506 63605: gap of unknown length
 63606 66610: contig of 3005 bp in length
 66611 72030: gap of unknown length
 72031 72130: contig of 5320 bp in length
 72131 75957: gap of unknown length
 75958 76057: gap of unknown length
 76058 79361: contig of 3304 bp in length
 79362 79461: gap of unknown length
 79462 84332: contig of 4861 bp in length
 84333 84482: gap of unknown length
 84483 91488: contig of 7066 bp in length
 91489 91588: gap of unknown length
 91589 99395: contig of 7807 bp in length.
 Location/Qualifiers
 1. 99395
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-8E2"
 BASE COUNT 26409 a 18452 c 18712 g 27333 t 8489 others
 ORIGIN
 Query Match 1.2%; Score 24; DB 2; Length 99395;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 591 CCTGTGACATCATCCACCAAGATG 614
 |||

Db 45144 CCTGTGACATATCCACAGATG 45121

RESULT 25
AC116944/c
LOCUS AC116944 131969 bp DNA linear HTG 14-JUL-2002
DEFINITION Tetraodon nigroviridis clone G5TNB-15B16, WORKING DRAFT SEQUENCE, 3 ordered pieces.

ACCESSION AC116944
VERSION AC116944.2 GI:21389263
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis

REFERENCE
AUTHORS Ahter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granter, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Latic, P., Lee, L.S.-O., Legaspi, R., Maduro, O.L., Maduro, V.B., Margulies, E.H., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Paquinigan, C., Pearson, R., Portnoy, M.E., Prasad, A., Schneider, M.G., Stantirlop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurguev, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.H. and Green, E.D.
NISC Comparative Sequencing Initiative

TITLE Unpublished
JOURNAL 2 (bases 1 to 131969)
REFERENCE Direct Submission
AUTHORS Green, E.D.
JOURNAL Submitted (03-APR-2002) NIH Intramural Sequencing Center, 8717
TITLE Government Circle, Gaithersburg, MD 20877, USA
JOURNAL 3 (bases 1 to 131969)
AUTHORS Green, E.D.
JOURNAL Direct Submission
TITLE Submitted (14-JUL-2002) NIH Intramural Sequencing Center, 8717
AUTHORS Government Circle, Gaithersburg, MD 20877, USA
JOURNAL On Jun 11, 2002 this sequence version replaced gi:19909418.
COMMENT ----- Genome Center

Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc-zoo@nigrl.nih.gov
Project Information
Center project name: cvt
Center clone name: 015B16

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 131566 bases at least Q40
Consensus quality: 131709 bases at least Q30
Consensus quality: 131757 bases at least Q20
Insert size: 135000; agarose-fp
Insert size: 131769; sum-of-contigs
Quality coverage: 10.24x in Q20 bases; agarose-fp
Quality coverage: 10.49x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 64284: contig of 64284 bp in length
* 64285 64384: gap of unknown length
* 64385 89965: contig of 25581 bp in length
* 89966 90065: gap of unknown length
* 90066 131969: contig of 41904 bp in length.
* Location/Qualifiers
1. 131969
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="G5TNB-15B16"
/clone_lib="G5TNB"
/note="Genoscope designation: COAB015B16"
1. 64284
/note="assembly-fragment
missing Sp6 clone end on 5' end of insert"
64385. 89965
/note="assembly-fragment"
90066. 131969
/note="assembly-fragment
clone_end:T7
vector_side:right"

BASE COUNT 36937 a 29847 c 29144 g 35841 t 200 others
ORIGIN

Query Match 1.2%: Score 24; DB 2; Length 131969;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATATAGAAAAA 2032
Db 102320 CCAATATAGAAAAA 102297

RESULT 26
AC010808 143655 bp DNA linear HTG 22-APR-2000
LOCUS AC010808
DEFINITION Homo sapiens clone RP11-3C7, WORKING DRAFT SEQUENCE, 3 unordered pieces.
ACCESSION AC010808
VERSION AC010808.4 GI:7637259
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Unpublished
JOURNAL 2 (bases 1 to 143655)
JOURNAL Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Balmain, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckgeater, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArnell, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Horton, L., Galagan, J., Gardy, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kahn, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, D., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Turrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

TITLE
JOURNAL
COMMENT

Wyman, D., Ye, W.-J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 22, 2000 this sequence version replaced gi:6454022.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L2713

Center clone name: 3_C7

Sequencing vector: M13: M77815; 100% of reads

Chemistry: Dye-terminator BigDye, 100% of reads

Assembly program: Phrap version 0.960731

Consensus quality: 125929 bases at least Q40

Consensus quality: 135995 bases at least Q30

Consensus quality: 141226 bases at least Q20

Insert size: 14000; agarose-gel

Insert size: 14345; sum-of-contigs

Quality coverage: 5.2 in Q20 bases; agarose-gel

Quality coverage: 5.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently

consists of 3 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 27688: contig of 27688 bp in length

* 27689 27788: gap of 100 bp

* 27789 74810: contig of 47022 bp in length

* 74811 74910: gap of 100 bp

* 74911 143655: contig of 68745 bp in length.

Location/Qualifiers

1. 143655

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="RP11-3C7"

/clone_1lb="RP11-11 Human Male BAC"

1. 27688

/note="assembly-fragment"

27789. 74810

/note="assembly-fragment"

clone_end:SP6

vector_side:left"

74911.143655

/note="assembly-fragment"

clone_end:77

vector_side:left"

BASE COUNT 45361 a 25207 c 25146 g 47741 t 200 others

Query Match 1.2% Score 24; DB 2; Length 143655;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2008 CCCAATAGCAAAAAA 2031
|||||
Db 110631 CCCAATAGCAAAAAA 110654

RESULT 27
AC073958/c
LOCUS
DEFINITION Homo sapiens BAC clone RP11-533L22 from 7, complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC073958
AC073958.4 GI:13162554
HTG.

Homo sapiens.

Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 168991)

Sullivan, J. E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

98063792

9847074

2 (bases 1 to 168991)

Goya, E., Maupin, R. and Garrett, J.

The sequence of Homo sapiens BAC clone RP11-533L22

Unpublished

3 (bases 1 to 168991)

Waterston, R. H.

Submitted (07-JUN-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 168991)

Waterston, R. H.

Submitted (28-FEB-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 168991)

Waterston, R.

Submitted (09-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

6 (bases 1 to 168991)

Waterston, R.

Submitted (10-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Feb 28, 2001 this sequence version replaced gi:9838250.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: sapiens@wustl.wustl.edu

----- Summary Statistics

Center project name: H_NH0533L22

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and

sequencing collaboration between the NHRI Chromosome 7 Mapping

Project (Eric D. Green, Director), John D. McPherson in the

Department of Genetics (Washington University), and the Washington

University Genome Sequencing Center. For additional information

about the map position of this sequence, see

http://www.nhgri.nih.gov/DIR/STB/CHRT, send

mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatem, M., Catanesse, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-204G8, 200 bp overlap.
Actual start of this clone is at base position 1 of RP11-533122;
actual end is at base position 6126 of RP11-204G8.

FEATURES

Source

1..168991
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-533122"
/clone_lib="RPCT-11"
1464..1580
/rpt_family="L1"
1676..1807
/rpt_family="Alu"
1854..2111
/rpt_family="Alu"
3316..3410
/rpt_family="MTR"
4346..4605
/rpt_family="L1"
4608..5284
/rpt_family="L1"
5305..5377
/rpt_family="L1"
5367..5627
/rpt_family="L1"
5619..6477
/rpt_family="L1"
6517..8268
/rpt_family="L1"
8372..8459
/rpt_family="A-rich"
9222..9465
/rpt_family="MaLR"
9907..10248
/rpt_family="ERV1"
10292..10426
/rpt_family="MTR"
10462..10754
/rpt_family="MaLR"
11266..11330
/rpt_family="GA-rich"
12412..12472
/rpt_family="MERL_type"
12731..13304
/rpt_family="ERV1"
13579..14191
/rpt_family="ERV1"
14271..14798
/rpt_family="ERV1"
14826..14914
/rpt_family="MTR"
14939..14966
/rpt_family="AT-rich"
14976..15080
/rpt_family="(TATATG)n"
15081..15377
/rpt_family="Alu"
15382..15803
/rpt_family="MERL_type"
15891..15919
/rpt_family="(A)n"

repeat_region 15960..16333
/rpt_family="Alu"
repeat_region 17258..17302
/rpt_family="(TCTA)n"
repeat_region 17322..17536
/rpt_family="(TA)n"
repeat_region 18033..18101
/rpt_family="L2"
repeat_region 18594..18680
/rpt_family="(TATATG)n"
repeat_region 20418..20469
/rpt_family="AT-rich"
repeat_region 20525..20552
/rpt_family="AT-rich"
repeat_region 20588..20609
/rpt_family="AT-rich"
repeat_region 21182..21440
/rpt_family="Alu"
repeat_region 21680..21858
/rpt_family="MERL_type"
repeat_region 22004..22317
/rpt_family="Alu"
repeat_region 22729..22771
/rpt_family="AT-rich"
repeat_region 22952..22974
/rpt_family="(CAAA)n"
repeat_region 25964..25995
/rpt_family="AT-rich"
repeat_region 27007..27028
/rpt_family="(T)n"
repeat_region 28716..28765
/rpt_family="AT-rich"
repeat_region 29559..29966
/rpt_family="MaLR"
repeat_region 31352..31791
/rpt_family="MaLR"
repeat_region 31976..32469
/rpt_family="MaLR"
repeat_region 32690..32736
/rpt_family="(CA)n"
repeat_region 33448..33625
/rpt_family="L2"
repeat_region 33912..33455
/rpt_family="L1"
repeat_region 34456..34747
/rpt_family="Alu"
repeat_region 34748..35738
/rpt_family="L1"
repeat_region 35712..36285
/rpt_family="L1"
repeat_region 36350..36604
/rpt_family="L1"

Query Match 1.2% Score 24; DB 9; Length 168991;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2008 CCCAATATAGAAAAA 2031
|||||
DB 151944 CCCAATATAGAAAAA 151921

RESULT 28
AC072055 181557 bp DNA linear HTG 23-SEP-2000
LOCUS Homo sapiens chromosome 7 clone RP11-48C19, WORKING DRAFT SEQUENCE,
DEFINITION 4 unordered pieces.
ACCESSION AC072055
VERSION AC072055.3 GI:10260931
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 181557)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 181557)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (07-JUN-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Sep 23, 2000 this sequence version replaced gi:9638044.

COMMENT ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Project Information -----
 Center project name: H_NH0048C19
 Summary Statistics -----
 Sequencing vector: M13; 73%
 Sequencing vector: plasmid; 27%
 Chemistry: Dye-terminator Big Dye; 73% of reads
 Chemistry: Dye-terminator Big Dye; 27% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 180582 bases at least Q40
 Consensus quality: 181076 bases at least Q30
 Consensus quality: 181385 bases at least Q20
 Insert size: 163000; agarose-fp
 Insert size: 183250; sum-of-contigs
 Quality coverage: 7.87 in Q20 bases; agarose-fp
 Quality coverage: 8.14 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 20897: contig of 20897 bp in length
 * 20898 20997: gap of unknown length
 * 20998 50690: contig of 29693 bp in length
 * 50691 50790: gap of unknown length
 * 50791 98248: contig of 47458 bp in length
 * 98249 98348: gap of unknown length
 * 98349 181557: contig of 83209 bp in length.
 Location/Qualifiers
 1. 181557
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /clone="RP11-48C19"
 1. 20897
 /note="assembly_name:Contig12"
 20998. 50690
 /note="assembly_name:Contig13"
 50791. 98248
 /note="assembly_name:Contig14
 clone_end:SP6
 vector_side:left"
 98349. 181557
 /note="assembly_name:Contig15
 clone_end:T7
 vector_side:right"
 BASE COUNT 59633 a 31324 c 31034 g 59262 t 304 others
 ORIGIN
 Query Match 1.2%; Score 24; DB 2; Length 181557;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2008 CCCAATACAGAAAAA 2031
 Db 114989 CCCAATACAGAAAAA 115012

RESULT 29
 AC116046
 LOCUS
 DEFINITION
 AC116046 191639 bp DNA linear HTG 23-MAR-2002
 ordered pieces.
 AC116046
 VERSION
 AC116046.1 GI:19697485
 HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE
 Papiu hamadryas
 Papiu hamadryas
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopitheciinae; Papiu.
 1 (bases 1 to 191639)
 Birren, B., Linton, L., Nusbaum, C., and Lander, E.
 2 (bases 1 to 191639)
 Unpublished

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

1 (bases 1 to 191639)
 Birren, B., Linton, L., Nusbaum, C., and Lander, E.
 2 (bases 1 to 191639)
 Unpublished

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Bouknight, B., Brown, A., Camarata, J., Campiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S.,
 Fard, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hags, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,
 McCarthy, M., McKernan, P., McKernan, K., Melirim, J., Meneses, L.,
 Mihov, T., Mlenga, V., Murphy, T., Naylor, T., Nguyen, C., Nicol, R.,
 Nordu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Piere, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
 Seaman, S., Severi, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Triggillo, T., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, J., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (23-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center -----
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information -----
 Center project name: L12531
 Center clone name: 343_P-17
 Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 189809 bases at least Q40
 Consensus quality: 190706 bases at least Q30
 Consensus quality: 191020 bases at least Q20
 Insert size: 192000; agarose-fp
 Insert size: 191339; sum-of-contigs
 Quality coverage: 10.7 in Q20 bases; agarose-fp
 Quality coverage: 10.7 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 40171: contig of 40171 bp in length
 * 40172 40271: gap of 100 bp
 * 40272 70180: contig of 29909 bp in length
 * 70181 70280: gap of 100 bp
 * 70281 156363: contig of 86083 bp in length
 * 156364 156463: gap of 100 bp
 * 156464 191639: contig of 35176 bp in length.
 Location/Qualifiers
 1..191639
 /organism="Papio hamadryas"
 /db_xref="taxon:9557"
 /clone="RP41-343P17"
 /clone_lib="RPCI-41 Male Baboon BAC"
 1..40171
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 40272..70180
 /note="assembly_fragment"
 70281..156363
 /note="assembly_fragment"
 156464..191639
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"
 BASE COUNT 55938 a 40368 c 42183 g 52844 t 306 others
 ORIGIN

Query Match 1.2%; Score 24; DB 2; Length 191639;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCNATAAGAAAAA 2032
 Db 43630 CCAATAGAAAAA 43653

RESULT 30
 AF176841 1740 bp DNA linear ROD 22-SEP-1999
 LOCUS
 DEFINITION Mus musculus intestine N-acetylglucosamine 6-O-sulfotransferase
 (I-GlcNAc-6-ST) gene, complete cds.
 ACCESSION AF176841
 VERSION AF176841.1 GI:5917711
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1740)
 AUTHORS Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
 TITLE Cloning and characterization of a mammalian
 N-acetylglucosamine-6-sulfotransferase that is highly restricted to
 intestinal tissue

JOURNAL Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
 MEDLINE 99423499
 PUBMED 10491328
 REFERENCE 2 (bases 1 to 1740)
 AUTHORS Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
 TITLE Direct Submission
 JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Avenue, Palo Alto, CA 94304, USA
 FEATURES
 1..1740
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57Bl/6"

/db_xref="taxon:10090"
 /chromosome="8"
 /map="8B1"
 /tissue_type="intestine"
 <405..>1592
 /gene="I-GlcNAc-6-ST"
 <405..>1592
 /gene="I-GlcNAc-6-ST"
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 405..1592
 /gene="I-GlcNAc-6-ST"
 /note="possibly involved in biosynthesis of mucosal
 ligands for L-selectin"
 /codon_start=1
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 /protein_id="AAD56003.1"
 /db_xref="GI:5917712"

CDS

BASE COUNT 365 a 513 c 491 g 371 t
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 1740;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 704 GTGCTCAGAGAGGCGCTCTT 726
 Db 918 GTGCTCAGAGAGGCGCTCTT 940

RESULT 31
 AF176840 1989 bp mRNA linear ROD 22-SEP-1999
 LOCUS
 DEFINITION Mus musculus intestine N-acetylglucosamine 6-O-sulfotransferase
 (I-GlcNAc-6-ST) mRNA, complete cds.
 ACCESSION AF176840
 VERSION AF176840.1 GI:5917709
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1989)
 AUTHORS Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
 TITLE Cloning and characterization of a mammalian
 N-acetylglucosamine-6-sulfotransferase that is highly restricted to
 intestinal tissue

JOURNAL Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
 MEDLINE 99423499
 PUBMED 10491328
 REFERENCE 2 (bases 1 to 1989)
 AUTHORS Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.
 TITLE Direct Submission
 JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Avenue, Palo Alto, CA 94304, USA
 FEATURES
 1..1989
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57Bl/6"
 /db_xref="taxon:10090"
 /chromosome="8"
 /map="8B1"
 /tissue_type="intestine"
 1..1989
 /gene="I-GlcNAc-6-ST"
 107..1294

gene
 CDS

```

/gene="1-G1CNA6-6-ST"
/possibly_involved_in_biosynthesis_of_mucosal
ligands_for_L-selectin"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD56002.1"
/db_xref="GI:5917710"
/translational="MRP.PRESSTVWMLSLWVQGIILFVLSRQVSSPAGIGEEVHVYLVLSRQSSSPVGGQLFSQHPDPVFLIMPAMWMTLSGGSA PALHMAVROLIRSPFLC DMDVEDAIVIPWRRIISDLPOMAVRALSCEPVGCAPFRGNISSEEVCKPLCATPFEL AQEACSSSHYLVLEKVEFFNFQIVYPLISDPALUKRVHLRDPRAVLRSEOTAKAL ARDQIVLGVNGVWEADPLRVAVNECRSHVRAEALAPPELDRVLYVEDD ARDPLVIRIELYAFRTGGLTLPQLOTWVHNIHGGSPGARRAEFTLDRALVSQAWR HULPEFAKIRVQELCGALQIOLIGRVSHELEFDLSLDLILPGMDSFWASSTERQK PES"
BASE COUNT      415 a      586 c      550 g      438 t
ORIGIN
Query Match      1.18; Score 23; DB 10; Length 1989;
Best Local Similarity 100.0%; Prid. No. 0.38;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      704      GTGGTCAGAGAGGTGGCGTCTT 726
Db      620      GTGCTCAAGAGAGGTGGCGTCTT 642
|||||
RESULT 32
LOCUS      BC011365      2002 bp      mRNA      linear      PRI 30-JUL-2001
DEFINITION      Homo sapiens, clone MGC:17245 IMAGE:4182593, mRNA, complete cds.
ACCESSION      BC011365
VERSION      BC011365.1      GI:15030211
KEYWORDS      MGC.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 2002)
Strausberg, R.
Direct Submission
Submitted (25-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Huiyik, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Project: 12 Row: 1 Column: 24
This clone was selected for full length sequencing because it
passed the following selection criteria: hexamer frequency ORF
analysis. Similarity but not identity to protein.
Location/Qualifiers
1..2002
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:17245 IMAGE:4182593"
/tissue_type="Brain, glioblastoma with EGFR amplification"
/clone_id="NCI_CGAP_Btn64"
FEATURES
SOURCE

```

```

CDS
      /lab_host="DH10B"
      /note="Vector: pCMV-Sport6"
      77..631
        /codon_start=1
        /product="Unknown (protein for MGC:17245)"
        /protein_id="AAH1365.1"
        /db_xref="GI:15030212"
        /translation="MARPDPGSGSLTDEALSYSDADPSMKDPLLOCTMLRKPKDKK
        SLDPYVLGLMTLQKCDPFLMKRSLYFLFADVENDNDPKEDERIALMRKLTLEIT
        HNMGTEDDETSTHGNSDSPRGFGHIIAIPVDYSACKREELSVKFKRPDDSKMG
        LAFIQDPDGWIELNPNKATLM"
BASE COUNT      609 a      383 c      401 g      609 t
ORIGIN
Query Match      1.1%; Score 23; DB 9; Length 2002;
Best Local Similarity 100.0%; Pident NO. 0.38;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2010 CAATATAGAAAAAAAAAAAAA 2032
|||||
Db 1975 CAATATAGAAAAAAAAAAAAA 1997

RESULT 33
LOCUS      BC013482          2459 bp      mRNA      linear      ROD 07-Aug-2002
DEFINITION Mus musculus, clone MGC:18906 IMAGE:4240772, mRNA, complete cds.
ACCESSION   BC013482
VERSION     BC013482.1 GI:15488669
KEYWORDS    MGC.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 2459)
AUTHORS     Strausberg,R
TITLE       Direct Submission
JOURNAL     Submitted (04-SEP-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: qcqpbs@email.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            CDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www.sngc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxill.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAX Plate: 25 Row: h Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, similarity but not identity to protein.
Location/Qualifiers
1..2459
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="EVE/N"
/clone="MGC:18906 IMAGE:4240772"
/tissue_type="kidney, normal, 5 month old male mouse."
/clone_id="NCI_CGAP_Kid14"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
42..1013
/codon_start=1
/product="Unknown (protein for MGC:18906)"

```

/db_xref="GI:1548670"
 /translation="MNSVSPRVVNDGHPFIPALGCTVDPKVPKDELIKATKAIADT
 GFRHDSAYLVQIEEYGOAIRSKIEDGTVKREDIYFTSKLWSTFHPRELVSCLEKT
 LKNAOLDVVDLYIIFHPALOGDGLFEPDEHGRKLAEAVDLDPTWAMEKCDAGLA
 KSTGSGNMFPEOLETILNKPSLKTAPYCNQVECHILNOSMDLYCKSKDILLYVCT
 LGSSDRKINWQKSPVLDDPVLICAMANKYOTPALAIRYOLGRIYLTTSFEKER
 IKEMKVEFPELASEDMKVLDDGHRNLRNTASYFDHPHPEFIDEY"

BASE COUNT 775 a 423 c 531 g 730 t
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 2459;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAAAAA 2032
 DB 2436 CAAATAGAAAAA 2458

RESULT 34
 BC026628 2488 bp mRNA linear ROD 07-AUG-2002
 LOCUS Mus musculus, expressed sequence AWS57061, clone MGC:37449
 DEFINITION IMAGE:4983237, mRNA, complete cds.
 ACCESSION BC026628
 VERSION BC026628.1 GI:20071841
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 Strausberg, R.

REFERENCE Direct Submission
 TITLE Submitted (02-APR-2002) National Institutes of Health, Mammalian
 JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 COMMENT Contact: MGC help desk

Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) medepaxil.stanford.edu
 Dickson, M., Schmutz, J., Grumwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAX Plate: 59 Row: n Column: 3
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES Location/Qualifiers

source 1..2488
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="MGC:37449 IMAGE:4983237"
 /tissue_type="Colon, normal, 5 month old male mouse."
 /clone_lib="NCI_CGAP_C024"
 /lab_host="DH10B"
 /note="Vector: PCMV-SPORT6"
 41..1012
 /codon_start=1
 /product="expressed sequence AWS57061"
 /protein_id="AAH26628.1"
 /db_xref="GI:20071842"

BASE COUNT 805 a 422 c 531 g 730 t
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 2488;
 Best Local Similarity 100.0%; Pred. No. 0.39;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAAAAA 2032
 DB 2435 CAAATAGAAAAA 2457

RESULT 35
 E12442/c 2660 bp DNA linear PAT 27-APR-1998
 LOCUS DNA encoding inter site of restriction enzyme between PstI and SmaI
 DEFINITION which contain O-acetylhomoserine sulphydrylase.
 ACCESSION E12442
 VERSION E12442.1 GI:3251275
 KEYWORDS JP 1996336391-A/2.
 SOURCE Acromonium chrysoeum.
 ORGANISM Acromonium chrysoeum.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreales; Hypocreaceae; mitosporic Hypocreaceae; Acromonium.
 Matsuda, A. and Muramatsu, S.
 Patent: JP 1996336391-A 2 24-DEC-1996;
 ASAHIT CHEM IND CO LTD
 OS Acromonium chrysoeum
 PN JP 1996336391-A/2
 PD 24-DEC-1996
 PF 13-JUN-1995 JP 1995145866
 PI MATSUDA AKIO, MURAMATSU SHUJI
 PC C12N15/09, C07H21/04, C12N1/15//C12N9/88, (C12N1/15, C12R1:645),
 PC (C12N9/88,
 PC C12R1:645);
 CC strandedness: Double;
 CC topology: linear;
 CC hypothetical: No;
 CC anti-sense: No;

key location/Qualifiers
 FH source 1..2860
 FH /organism="Acromonium chrysoeum" FT
 FT /strain="IS-5"
 FT 5'UTR 1..448
 FT mat-peptide 449..495
 FT /product="inter site of restriction enzyme
 FT between PstI and
 FT SmaI which contain O-acetylhomoserine FT
 FT sulphydrylase"
 FT mat-peptide 647..701
 FT /product="inter site of restriction enzyme
 FT between PstI and
 FT SmaI which contain O-acetylhomoserine FT
 FT sulphydrylase"
 FT mat-peptide 1316..1914
 FT /product="inter site of restriction enzyme
 FT between PstI and
 FT SmaI which contain O-acetylhomoserine FT

[illegible]

| | |
|------------|--|
| RESULT 37 | |
| LOCUS | AC116989 |
| DEFINITION | AC116989 55470 bp DNA linear HTG 24 -MAY-2002 |
| ACCESSION | AC116989 |
| VERSION | AC116989.1 |
| KEYWORDS | AXI, *** SEQUENCING IN PROGRESS ***, in ordered pieces. |
| SOURCE | HTG: HTGS_PHRASE2. |
| ORGANISM | Dictyostelium discoideum. |
| REFERENCE | Dictyostelium discoideum. |
| AUTHORS | Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium. |
| | 1 (bases 1 to 55470) |
| | Gloeckner,G., Eichinger,L., Szafrański,K., Pachepat,J., Dear,P., |
| | Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K., |
| | Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and |
| | Noegel,A.A. |
| | Sequence and Analysis of Chromosome 2 of Dictyostelium |
| TITLE | Unpublished |
| JOURNAL | The Dictyostelium Genome Sequencing Consortium |
| REMARK | 2 (bases 1 to 55470) |
| REFERENCE | Baumgart,C. |
| AUTHORS | Direct Submission |
| JOURNAL | Submitted (05-APR-2002) Genome Analysis, Institute of Molecular |
| | Biotechnology, Beutenbergstr. 11, Jena 07745, Germany |
| | 3 (bases 1 to 55470) |
| | Baumgart,C. |
| | Direct Submission |
| | Submitted (24-MAY-2002) Genome Analysis, Institute of Molecular |
| | Biotechnology, Beutenbergstr. 11, Jena 07745, Germany |
| | CDS predictions from GenEd may contain errors. Further information |
| | is available from IMB Jena, Department of Genome Analysis |
| | (http://genome.imb-jena.de/dictyostelium/) |
| | and the University Cologne, Institute for Biochemistry I |
| | (http://www.uni-koeln.de/dictyostelium/project.shtml |
| | Funding |
| COMMENT | Agency: Deutsche Forschungsgemeinschaft (DFG). |
| | * NOTE: This is a 'working draft' sequence. |
| | * This sequence will be replaced |
| | * by the finished sequence as soon as it is available and |
| | * the accession number will be preserved. |
| FEATURES | location/Qualifiers |
| source | 1..55470 |
| | /organism="Dictyostelium discoideum" |
| | /strain="AX4" |
| | /db_xref="taxon:44689" |
| | /chromosome="2" |
| | /map="3542416-3597884" |
| | 2458..6183 |
| | /note="ORF ID:dd_03152" |
| | /codon_start=1 |
| | /product="hypothetical protein" |
| | /protein_id="AA043736.1" |
| | /db_xref="gi:2116119" |
| | /translation="MEDEETKRLNERRRILNEERKKQEDDAKEDDERKAREEERKYV |
| | RKEKREKEREERAOCKLELEYKRTFEEREOREERERKRIEEOCKDEEKRKR |
| | QOEERKEREDEDEKRRKILPEERKALKEEEDENSQVSLISOKRRRIARQ |
| | KISLIALSDPSNSQVQPOPTQPTKPIYSLITLDELKEVIDIADVGGFRLYDMOSP |
| | ASISMAEYLVIDVNGATIRASKREPIYNGASLEKRLSARMAQSFTLGGQDTAK |
| | IITDPSFLIDILFMSITSNPPPLINMHLFDVLQVSYSRKELLVYICNNKDTIL |
| | PSLQHLNSETVYDLKMAEIENTQNETTSSSSGSAASFEFTGSSSSSSSS |
| | NNNNNNNNKREPMWTMDLPYLLRLNETSINENYVDQIDLEIDNIORLMEFLVK |
| | YPSKVAHLGCGFSNAVNRGRLALVHYTEPFGSSMSLICTLVDSSVTQGL |
| | ITNNSSSSSSLSSSSSSSLSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS |
| | ELSKIKLSIETITTKKDSQELFSYRRCALHINSYKANTSKIDPREIMSGILKCI |
| | DIFREYTNVYHGMIGCILPLIQRLOGSDDEPMCHLKDSQFSRVYNTLSGNNN |
| | SELQIINNSNNNNNDSSNNNNNNNNNNNNINSISGAKSLALFNYSKDNQVYIVE |
| | QOQSEYVIGAPAPMAGSLSGGGKRIKRYKFSFTDEDSINADENNISNTNGDYK |
| | LAIGDELMPQPTSTPAPATSTEDQSDKSLNENTVYSKRFATLSTFSISSSS |
| | SSNTSSTSTTTTKTNGTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS |
| | LTPDQEAIRKSOIKRLKEREKREKKLAKRLKEREKREKSTSTSSSTLKNSS |
| | GSVLTPTTKTKLSSSSSSSSSSSSSSSVLTSTSTIAATTTTTPTPTATPTTTTT |
| | TTTTTTTTTSLDSDSTSSSDCSQSDVSSSEISKSLINDNSQPLDASVENIESIE |
| | QSVVVVYHNESSNNGSSNTECNNSINDESSNTPTSTTTATTKKRVNHNVGHI |

```

CDS
        join(17865..17877,17959..18215)
        /note="ORF_ID:dd_01367"
        /codon_start=1
        /product="hypothetical protein"
        /protein_id="AA043745.1"
        /db_xref="GI:21166128"
        /translation="MTLTIASISIGNKAIKSKSNPFSSLSQSSNSIQCGCGGGG
        NSLIGTVGLGVAVGGLVGTGIVGVGVNNGVGGLLSGPCCGH"
        complement(join(18632..18888,18961..18973))
        /note="ORF_ID:dd_01385"
        /codon_start=1
        /product="hypothetical protein"
        /protein_id="AA043746.1"
        /db_xref="GI:21166129"
        /translation="MTLTIASISIGNKAIKSKSNPFSSLSQSSNSIQCGCGGGG
        SPLIGTVGLGVAVGGLVGTGIVGVGVNNGVGGLLSGPCCGH"
        join(19666..19678,19759..20015)
        /note="ORF_ID:dd_01382"
        /codon_start=1
        /product="hypothetical protein"
        /protein_id="AA043747.1"
        /db_xref="GI:21166130"
        /translation="MTLTIASISIGNKAIKSKSNPFSSLSQSSNSIQCGCGGGG
        SPLIGTVGLGVAVGGLVGTGIVGVGVNNGVGGLLSGPCCGH"
        complement(join(20170..20736,20806..20864,20943..21345))
        /note="ORF_ID:dd_01379"
        /codon_start=1
        /product="putative oxidoreductase"
        /protein_id="AA043748.1"
        /db_xref="GI:21166131"
        /translation="MGVITDISLITNTNIGSIIILFYPIGLFLNKKKKTNPFPFA
        IYITRASSGIGTGIALEAKRRKNNSLVAGTGTNRIKLEPTEKCEICGVQVEEST
        DVTBKEKIDMLIKFDNKYIKIDILIANAGTETILPKLDEITELIVANTVNIAGTIN
        TVLPEKLVIFESRSGQVLWMSITPYEDYVAVGSSSKGYIKSGGLIRNGLRAGVG
        VSAVIRGIVPIPLVDSDYKIDIPMLPILPTTDKASKYIVDGISRDAITSPPLIYCS
        HFVAITPIPLDVAITNFISNLSLKYPDYKPYTHSHSHSQYTNSKREDDPSSPAANMS
        NSORKKEA"
        21759..22802
        /note="ORF_ID:dd_01376"
        /codon_start=1
        /product="cysteine-rich, acidic integral membrane protein
        precursor"
        /protein_id="AA043749.1"

Query Match      1.1%  Score 23:  DB 2:  Length 55470;
Best Local Similarity 100.0%;  Pred. NO. 0.45;
Matches 23:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:

QY      2010  CAAATAGAGAAAAA 2032
Db      53222  CAAATAGAGAAAAA 53244

RESULT 38
AC103270/LOCUS      56342 bp  DNA      linear      HTG 13-JUL-2002
DEFINITION      Rattus norvegicus clone CH230-228B1, *** SQUENCING IN PROGRESS
AC103270/VERSION      AC103270.3  GI:21731258

```

KEYWORDS

HTG: HTGS, PHASE1.

SOURCE

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 56342)

Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C., Albrooks,S.L., Amaralung,H.C., Are,J.R., Ayale,M., Banks,T., Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D., Bouck,J., Bowie,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P., Butay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Daborne,S.R., David,R., Davila,M.L., Davis,C., Davy-carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huik,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochvil,A., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,M., Louisged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,K., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Matlindale,A., Martinez,E., Massey,E., Mawney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabhat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S., Ogih,M., Okunomi,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Uemari,K., Vasquez,L., Vera,Y., Villalob,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE

Unpublished

JOURNAL

2 (bases 1 to 56342)

REFERENCE

AUTHORS

Worley,K.C.

TITLE

Direct Submission

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 56342)

REFERENCE

AUTHORS

Worley,K.C.

TITLE

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 11, 2002 this sequence version replaced gi:17974782.

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G102
Center clone name: CH230-228B1
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.990329
Consensus quality: 9543 bases at least Q40
Consensus quality: 10073 bases at least Q30
Consensus quality: 10542 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1205: contig of 1205 bp in length
1206 1305: gap of unknown length
1306 2373: contig of 1068 bp in length
2374 2473: gap of unknown length
2474 3564: contig of 1091 bp in length
3565 4730: gap of unknown length
4731 4830: contig of 1066 bp in length
4831 6009: contig of 1179 bp in length
6010 6109: gap of unknown length
6110 7354: contig of 1245 bp in length
7355 7454: gap of unknown length
7455 8697: contig of 1243 bp in length
8698 8797: gap of unknown length
8798 9797: contig of 1000 bp in length
9798 9897: gap of unknown length
9898 10909: contig of 1012 bp in length
10910 11009: gap of unknown length
11010 12689: contig of 1680 bp in length
12690 12789: gap of unknown length
12790 14479: contig of 1690 bp in length
14480 14579: gap of unknown length
14580 15724: contig of 1145 bp in length
15725 15824: gap of unknown length
15825 16946: contig of 1122 bp in length
16947 17046: gap of unknown length
17047 18079: contig of 1033 bp in length
18080 18179: gap of unknown length
18180 19855: contig of 1676 bp in length
19856 19955: gap of unknown length
19956 20982: contig of 1027 bp in length
20983 21082: gap of unknown length
21083 22928: contig of 1846 bp in length
22929 23028: gap of unknown length
23029 24033: contig of 1005 bp in length
24034 24133: gap of unknown length
24134 25243: contig of 1110 bp in length
25244 25343: gap of unknown length
25344 27165: contig of 1822 bp in length
27166 27265: gap of unknown length
27266 28838: gap of unknown length
28839 28938: gap of unknown length
28939 30372: contig of 1434 bp in length
30373 30472: gap of unknown length
30473 32069: contig of 1597 bp in length
32070 32169: gap of unknown length
32170 34053: contig of 1884 bp in length
34054 34153: gap of unknown length
34154 35815: contig of 1662 bp in length
35816 35915: gap of unknown length
35916 37432: contig of 1517 bp in length
37433 37532: gap of unknown length
37533 39072: contig of 1540 bp in length
39073 39172: gap of unknown length
39173 40536: contig of 1367 bp in length
40537 40540: gap of unknown length
40540 42394: contig of 1755 bp in length
42395 42494: gap of unknown length

```

* 42495 44274: contig of 1780 bp in length
* 44275 44374: gap of unknown length
* 44375 46372: contig of 2198 bp in length
* 46373 46673: gap of unknown length
* 46673 47937: contig of 1265 bp in length
* 47938 48037: gap of unknown length
* 48038 49799: contig of 1762 bp in length
* 49800 49899: gap of unknown length
* 49900 51038: contig of 1139 bp in length
* 51039 51138: gap of unknown length
* 51139 56342: contig of 5204 bp in length.
Location/Qualifiers
1..56342
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-228B1"

BASE COUNT 15092 a 11063 c 11162 g 13872 t 5153 others
ORIGIN

Query Match 1.18; Score 23; DB 2; Length 56342;
Best Local Similarity 100.0%; Pred. No. 0.45;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2010 CAATTAAGAAAAA 2032
|||||
Db 6829 CAATTAAGAAAAA 6807

RESULT 39
AC087621 63739 bp DNA linear HTG 15-JAN-2001
LOCUS Homo sapiens chromosome 8 clone RP11-300E4 map 8, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
AC087621
AC087621.1 GI:12229396
HTG: HTGS-PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 63739)
Homo sapiens chromosome 8, clone RP11-300E4
2 (bases 1 to 63739)
Unpublished
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barnes,N., Bastien,V., Boguski,M., Bouckgeat,B., Brown,A.,
Camata,J., Campiano,A., Ciopele,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardua,S., Glade,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karas,A., Larocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
Margis,N., Mathews,C., McCarthy,M., McKean,P., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,C., Mhova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norb,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Plunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retter,S.,
Riebeck,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Sougez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Sudriamanian,A., Talamas,J., Testaye,S., Theodore,J.,
Travers,M., Travis,N., Triglio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zambek,L., Zimmer,A. and Zody,M.

TITLE Submitted (15-JAN-2001) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11683
Center clone name: 300_E4
NOTE: This record contains 79 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
701 800: gap of 100 bp
801 1491: contig of 691 bp in length
1492 1591: gap of 100 bp
1592 2304: contig of 713 bp in length
2305 2404: gap of 100 bp
2405 3109: contig of 705 bp in length
3110 3209: gap of 100 bp
3210 3926: contig of 717 bp in length
3927 4026: gap of 100 bp
4027 4745: contig of 719 bp in length
4746 4845: gap of 100 bp
4846 5573: contig of 728 bp in length
5574 5673: gap of 100 bp
5674 6393: contig of 722 bp in length
6396 6495: gap of 100 bp
6496 7204: contig of 709 bp in length
7205 7304: gap of 100 bp
7305 8014: contig of 710 bp in length
8015 8114: gap of 100 bp
8115 8828: contig of 714 bp in length
8829 8928: gap of 100 bp
8929 9622: contig of 694 bp in length
9623 9722: gap of 100 bp
9723 10431: contig of 709 bp in length
10432 10531: gap of 100 bp
10532 11330: contig of 699 bp in length
11331 11330: gap of 100 bp
11331 11870: contig of 640 bp in length
11971 12070: gap of 100 bp
12071 12823: contig of 753 bp in length
12824 12923: gap of 100 bp
12924 13630: contig of 707 bp in length
13631 13730: gap of 100 bp
13731 14448: contig of 718 bp in length
14449 14548: gap of 100 bp
14549 15262: contig of 714 bp in length
15263 15362: gap of 100 bp
15363 16076: contig of 714 bp in length
16077 16176: gap of 100 bp
16177 16880: contig of 704 bp in length
16881 16980: gap of 100 bp
16981 17177: contig of 737 bp in length
17178 17817: gap of 100 bp
17818 18522: contig of 705 bp in length
18523 18622: gap of 100 bp
18623 19343: contig of 721 bp in length
19344 19443: gap of 100 bp
19444 20161: contig of 718 bp in length
20162 20261: gap of 100 bp
20262 20983: contig of 722 bp in length
20984 21083: gap of 100 bp
21084 21810: contig of 727 bp in length
21811 21910: gap of 100 bp

```


* 21911 22637: contig of 727 bp in length
* 22638 22737: gap of 100 bp
* 22738 23439: contig of 702 bp in length
* 23440 23539: gap of 100 bp
* 23540 24247: contig of 708 bp in length
* 24248 24347: gap of 100 bp
* 24348 25051: contig of 704 bp in length
* 25052 25151: gap of 100 bp
* 25152 25851: contig of 700 bp in length
* 25852 25951: gap of 100 bp
* 25952 26675: contig of 724 bp in length
* 26676 26775: gap of 100 bp
* 26776 27470: contig of 695 bp in length
* 27471 27570: gap of 100 bp
* 27571 28264: contig of 694 bp in length
* 28265 28364: gap of 100 bp
* 28365 29085: contig of 721 bp in length
* 29086 29185: gap of 100 bp
* 29186 29892: contig of 707 bp in length
* 29893 29992: gap of 100 bp
* 29993 30703: contig of 711 bp in length
* 30704 30803: gap of 100 bp
* 30804 31509: contig of 706 bp in length
* 31510 31609: gap of 100 bp
* 31610 32297: contig of 688 bp in length
* 32298 32397: gap of 100 bp
* 32398 33104: contig of 707 bp in length
* 33105 33204: gap of 100 bp
* 33205 33905: contig of 701 bp in length
* 33906 34005: gap of 100 bp
* 34006 34708: contig of 703 bp in length
* 34709 34808: gap of 100 bp
* 34809 35513: contig of 705 bp in length
* 35514 35613: gap of 100 bp
* 35614 36345: contig of 732 bp in length
* 36346 36445: gap of 100 bp
* 36446 37174: contig of 729 bp in length
* 37175 37274: gap of 100 bp
* 37275 37987: contig of 713 bp in length
* 37988 38087: gap of 100 bp
* 38088 38800: contig of 713 bp in length
* 38801 38900: gap of 100 bp
* 38901 39598: contig of 698 bp in length
* 39599 39698: gap of 100 bp
* 39699 40403: contig of 705 bp in length
* 40404 40503: gap of 100 bp
* 40504 41221: contig of 718 bp in length
* 41222 41321: gap of 100 bp
* 41322 42007: contig of 686 bp in length
* 42008 42107: gap of 100 bp
* 42108 42794: contig of 687 bp in length
* 42795 42894: gap of 100 bp
* 42895 43612: contig of 718 bp in length
* 43613 43712: gap of 100 bp
* 43713 44428: contig of 717 bp in length
* 44430 44529: gap of 100 bp
* 44530 45239: contig of 710 bp in length
* 45240 45339: gap of 100 bp
* 45340 46049: contig of 710 bp in length
* 46050 46149: gap of 100 bp
* 46150 46864: contig of 715 bp in length
* 46865 46964: gap of 100 bp
* 46965 47674: contig of 710 bp in length
* 47675 47774: gap of 100 bp
* 47775 48465: contig of 691 bp in length
* 48466 48565: gap of 100 bp
* 48566 49271: contig of 706 bp in length
* 49272 49371: gap of 100 bp
* 49372 50018: contig of 647 bp in length
* 50019 50118: gap of 100 bp
* 50119 50839: contig of 721 bp in length
* 50840 50939: gap of 100 bp
* 50940 51664: contig of 725 bp in length

* 51665 51764: gap of 100 bp
* 51765 52477: contig of 713 bp in length
* 52478 52577: gap of 100 bp
* 52578 53281: contig of 704 bp in length
* 53282 53381: gap of 100 bp
* 53382 54081: contig of 700 bp in length
* 54082 54181: gap of 100 bp
* 54182 54877: contig of 696 bp in length
* 54878 54977: gap of 100 bp
* 54978 55682: contig of 705 bp in length
* 55683 55782: gap of 100 bp

Query Match 1.1% Score 23: DB 2: Length 63739;
Best Local Similarity 100.0% Fred. No. 0.46;

Matches 23: Conservative 0: Mismatches 0: Indels 0: Gaps 0;

QY 2010 CAATTAAGAAAAA 2032

Db 33706 CAATTAAGAAAAA 33728

RESULT 40

AC124305/c

LOCUS

DEFINITION

AC124305

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I27334

Center clone name: 173_H_16

* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

707: contig of 707 bp in length
1 708 807: gap of 100 bp
808 1544: contig of 737 bp in length
1545 1644: gap of 100 bp
1645 2355: contig of 711 bp in length
2356 2455: gap of 100 bp
2456 3193: contig of 744 bp in length
3200 3299: gap of 100 bp
3300 4006: contig of 707 bp in length
4007 4106: gap of 100 bp
4107 4838: contig of 732 bp in length
4839 4938: gap of 100 bp
4939 5676: contig of 738 bp in length
5677 5776: gap of 100 bp
5777 6430: contig of 714 bp in length
6431 6590: gap of 100 bp
6591 7307: contig of 717 bp in length
7308 7407: gap of 100 bp
7408 8136: contig of 729 bp in length
8137 8236: gap of 100 bp
8237 8975: contig of 739 bp in length
8976 9075: gap of 100 bp
9076 9817: contig of 742 bp in length
9818 9917: gap of 100 bp
9918 10665: contig of 748 bp in length
10666 10765: gap of 100 bp
10766 11510: contig of 745 bp in length
11511 11610: gap of 100 bp
11611 12353: contig of 743 bp in length
12354 12453: gap of 100 bp
12454 13195: contig of 742 bp in length
13196 13295: gap of 100 bp
13296 14021: contig of 726 bp in length
14022 14121: gap of 100 bp
14122 14839: contig of 718 bp in length
14840 14939: gap of 100 bp
14940 15646: contig of 707 bp in length
15647 15746: gap of 100 bp
15747 16482: contig of 736 bp in length
16483 16582: gap of 100 bp
16583 17280: contig of 698 bp in length
17281 17380: gap of 100 bp
17381 18109: contig of 729 bp in length
18110 18209: gap of 100 bp
18210 18949: contig of 740 bp in length
18950 19049: gap of 100 bp
19050 19783: contig of 734 bp in length
19784 19883: gap of 100 bp
19884 20602: contig of 719 bp in length
20603 20702: gap of 100 bp
20703 21439: contig of 737 bp in length
21440 21539: gap of 100 bp
21540 22271: contig of 732 bp in length
22272 22371: gap of 100 bp
22372 23075: contig of 704 bp in length
23076 23175: gap of 100 bp
23176 23875: contig of 700 bp in length
23876 23975: gap of 100 bp
23976 24708: contig of 733 bp in length
24709 24808: gap of 100 bp

24809 25543: contig of 735 bp in length
2544 25643: gap of 100 bp
25644 26383: contig of 740 bp in length
26384 26483: gap of 100 bp
26484 27219: contig of 726 bp in length
27220 27319: gap of 100 bp
27320 28060: contig of 741 bp in length
28061 28160: gap of 100 bp
28161 28896: contig of 736 bp in length
28897 28996: gap of 100 bp
28997 29734: contig of 738 bp in length
29735 29834: gap of 100 bp
29835 30568: contig of 734 bp in length
30569 30668: gap of 100 bp
30669 31405: contig of 737 bp in length
31406 31505: gap of 100 bp
31506 32230: contig of 725 bp in length
32231 32330: gap of 100 bp
32332 33057: contig of 727 bp in length
33058 33157: gap of 100 bp
33158 33892: contig of 735 bp in length
33893 33992: gap of 100 bp
33993 34726: contig of 734 bp in length
34727 34826: gap of 100 bp
34827 35528: contig of 702 bp in length
35529 35628: gap of 100 bp
35629 36361: contig of 733 bp in length
36362 36461: gap of 100 bp
36462 37200: contig of 739 bp in length
37201 37300: gap of 100 bp
37301 38028: contig of 728 bp in length
38029 38128: gap of 100 bp
38129 38865: contig of 737 bp in length
38866 38965: gap of 100 bp
38966 39705: contig of 740 bp in length
39706 39805: gap of 100 bp
39806 40538: contig of 733 bp in length
40539 40638: gap of 100 bp
40639 41354: contig of 716 bp in length
41355 41454: gap of 100 bp
41455 42181: contig of 727 bp in length
42182 42281: gap of 100 bp
42282 42999: contig of 718 bp in length
43000 43099: gap of 100 bp
43100 43833: contig of 734 bp in length
43834 43933: gap of 100 bp
43934 44667: contig of 734 bp in length
44668 44767: gap of 100 bp
44768 45472: contig of 705 bp in length
45473 45572: gap of 100 bp
45573 46309: contig of 737 bp in length
46310 46409: gap of 100 bp
46410 47110: contig of 701 bp in length
47111 47210: gap of 100 bp
47211 47918: contig of 708 bp in length
47919 48018: gap of 100 bp
48019 48746: contig of 728 bp in length
48747 48846: gap of 100 bp
48847 49582: contig of 736 bp in length
49583 49682: gap of 100 bp
49683 50413: contig of 737 bp in length
50420 50519: gap of 100 bp
50520 51224: contig of 705 bp in length
51225 51324: gap of 100 bp
51325 52061: contig of 737 bp in length
52062 52161: gap of 100 bp
52162 52873: contig of 712 bp in length
52874 52973: gap of 100 bp
52974 53688: contig of 715 bp in length
53689 53788: gap of 100 bp
53789 54521: contig of 733 bp in length
54522 54621: gap of 100 bp
54622 55327: contig of 706 bp in length

* 55328 55427: gap of 100 bp
 * 55428 56157: contig of 730 bp in length
 * 56158 56257: gap of 100 bp
 * 56258 56999: contig of 742 bp in length

Query Match 1.1%; Score 23; DB 2; Length 69523;
 Best Local Similarity 100.0%; Pred. No. 0.46;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAAGCAAAAAA 2032
 Db 40400 CAATAAGCAAAAAA 40378

RESULT 41
 AC083858 74718 bp DNA linear ROD 05-OCT-2001
 LOCUS Mus musculus chromosome 5 clone RP23-423A22 strain C57BL6/J,
 DEFINITION complete sequence.
 AC083858
 VERSION AC083858.3 GI:15963650
 KEYWORDS HTG.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 74718)
 AUTHORS Ahlter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Ho, S.-L., Idol, J.R., Karlins, E., Latic, P., Lee-Lim, S.-Q.,
 Legaspi, R., Maduro, Q.L., Maduro, V.B., Masello, C., Mastrian, S.D.,
 McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y.,
 Stantropop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
 Tsurgren, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Zhang, L., H. and Green, E.D.
 TITLE NISC Comparative Sequencing Initiative
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 74718)
 AUTHORS Green, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (04-OCT-2000) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 REFERENCE 3 (bases 1 to 74718)
 AUTHORS Green, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-2001) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 COMMENT On Oct 5, 2001 this sequence version replaced gi:12313754.
 Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc.mouse@nhgri.nih.gov
 Project Information
 Center project name: r9
 Center clone name: 423A22

This sequence was finished as follows unless otherwise noted:
 all regions were double-stranded, sequenced with an
 alternate chemistry, or covered by high quality data
 (i.e., phred quality >= 30); an attempt was made to resolve
 all sequencing problems, such as compressions and repeats;
 all regions were covered by at least one plasmid subclone
 or more than one M13 subclone; and the assembly was confirmed
 by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
 this clone unless otherwise noted. If there are overlapping
 clones, the overlaps are noted in the beginning and end of
 the features section.

FEATURES
 Location/Qualifiers
 1..74718

misc-feature

misc-feature

misc-feature

BASE COUNT 19836 a 17184 c 17496 g 20202 t

Query Match 1.1%; Score 23; DB 10; Length 74718;
 Best Local Similarity 100.0%; Pred. No. 0.46;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAAGCAAAAAA 2032
 Db 40963 CAATAAGCAAAAAA 40985

RESULT 42
 AC117427/c 83543 bp DNA linear PRI 25-MAY-2002
 LOCUS Homo sapiens 3 BAC RP11-6615 (Roswell Park Cancer Institute Human
 DEFINITION BAC library) complete sequence.
 AC117427
 VERSION AC117427.2 GI:21206085
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 83543)
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Selman, F.R., Allen, C.,
 Alstbrooks, S.L., Anarlungue, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbier, J., Benton, D., Blmage, K., Blankensbury, K., Bonnin, D.,
 Bouck, J., Bowls, S., Briley, M., Brown, E., Brown, M., Bryant, N.P.,
 Buha, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C.,
 Carion, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Day-Carroll, L., Dedetich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinu, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabor, J., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrill, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., He, X.,
 Hernandez, U., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
 Hollins, B., Homsl, F., Howard, S., Huber, J., Huylk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Lantry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtenberg, O., Liu, C., Liu, J., Liu, W.,
 Louisseg, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawlin, E., McLeod, M.P., Meador, M.,
 Mel, G., Mettler, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Moore, S., Morgan, M., Morris, T., Morris, S., Mosser, M., Neal, D.,

[illegible]

FEATURES

| | |
|---------------|--|
| URS | location/Qualifiers |
| source | 1. .83543 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon.9606" |
| | /chromosome="3" |
| | /clone="RP11-6615" |
| | /complement(1..2001) |
| | /note="overlaps bases 1..2001 of clone AC108698" |
| misc-feature | /function="clone overlap" |
| repeat_region | /complement(1..112) |
| repeat_region | /rpt_family="L1R24" |
| repeat_region | /complement(113..229) |
| repeat_region | /rpt_family="HERV23" |
| repeat_region | /complement(601..746) |
| repeat_region | /rpt_family="HERV118" |
| repeat_region | 1863..1888 |
| repeat_region | /rpt_family="(T)n" |
| repeat_region | /complement(3930..4376) |
| repeat_region | /rpt_family="MLR2B" |
| repeat_region | /complement(4380..5057) |
| repeat_region | /rpt_family="HERVL" |
| repeat_region | /complement(6785..7252) |
| repeat_region | /rpt_family="MLR2A" |
| repeat_region | /complement(7904..7992) |
| repeat_region | /rpt_family="HERV9" |
| repeat_region | 8323..8343 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 8493..9140 |
| repeat_region | /rpt_family="L1PA7" |
| repeat_region | 9228..9298 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 9652..9672 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 10809..10841 |
| repeat_region | /rpt_family="(T)n" |
| repeat_region | /complement(11441..11940) |
| repeat_region | /rpt_family="HERV23" |
| repeat_region | /complement(11941..12409) |
| repeat_region | /rpt_family="L1R24" |
| repeat_region | /complement(12808..13105) |
| repeat_region | /rpt_family="A1u3b" |
| repeat_region | /complement(13880..14322) |
| repeat_region | /rpt_family="L1R38" |
| repeat_region | /complement(14840..15069) |
| repeat_region | /rpt_family="L1R38" |
| STS | 15307..15506 |
| repeat_region | /standard_name="172140" |
| repeat_region | /complement(15347..15937) |
| repeat_region | /rpt_family="L1R49" |
| repeat_region | 16613..16941 |
| repeat_region | /rpt_family="MER67D" |
| repeat_region | 16980..17164 |
| repeat_region | /rpt_family="MER67D" |
| repeat_region | 17389..17423 |
| repeat_region | /rpt_family="(A)n" |
| repeat_region | 18145..19797 |
| repeat_region | /rpt_family="MER52A" |
| repeat_region | 21274..21319 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | /complement(21578..21951) |
| repeat_region | /rpt_family="MER67D" |
| repeat_region | 22410..22431 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 22615..22923 |
| repeat_region | /rpt_family="A1u9" |
| repeat_region | 24740..24790 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 25202..25256 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 25543..26656 |
| repeat_region | /rpt_family="L1PA13" |
| repeat_region | 26857..26884 |
| repeat_region | /rpt_family="(CA)n" |

```

repeat_region      26895..27733
                    /rpt_family="L1PA13"
repeat_region      28372..29543
                    /rpt_family="L1PA16"
repeat_region      29544..30552
                    /rpt_family="MER11A"
repeat_region      30553..30863
                    /rpt_family="L1PA16"
repeat_region      30864..30906
                    /rpt_family="L1PA16"
repeat_region      30907..32672
                    /rpt_family="L1PA16"
repeat_region      32673..33726
                    /rpt_family="MER11C"
repeat_region      33727..34480
                    /rpt_family="L1PA16"

```

```

Query Match      1.1%; Score 23; DB 9; Length 83543;
Best Local Similarity 100.0%; Pred.No. 0.46;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2010 CAATAGAGAAAAA 2032
Db 73623 CAATAGAGAAAAA 73601

```

```

RESULT 43
AL590368/c      92910 bp DNA linear PRI 15-NOV-2001
LOCUS      Human DNA sequence from clone RP11-3K19 on chromosome 9, complete
DEFINITION
ACCESSION      AL590368 AC044833
VERSION      AL590368.11 GI:16973067
KEYWORDS      HNG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 92910)
AUTHORS      Bates,K.
JOURNAL      Direct Submission
Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Nov 16, 2001 this sequence version replaced gi:16416219.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormBase; Information on the WormBase
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormbase This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-3K19 is from the library RPO1-11.1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/dacpac/home.htm
VECTOR: pBAC3.6

```

IMPORTANT: This sequence is not the entire insert of clone RP11-3K19 it may be shorter because we sequence overlapping

sections only once, except for a short overlap.
The true left end of clone RP11-477A7 is at 90911 in this sequence.
The true right end of clone RP11-112N13 is at 2000 in this sequence.

```

FEATURES
    source
        1..92910
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-3K19"
            /clone_11b="RPO1-11.1"
            /clone_11p="RPO1-11.1"

```

```

misc-feature
    /note="Single clone region. Sequence from reads from a
    short insert library derived from a single pUC clone.
    Restriction digest data confirm the assembly."
BASE COUNT      25801 a 18381 c 18458 g 30270 t
ORIGIN

```

```

Query Match      1.1%; Score 23; DB 9; Length 92910;
Best Local Similarity 100.0%; Pred.No. 0.47;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2010 CAATAGAGAAAAA 2032
Db 24900 CAATAGAGAAAAA 24878

```

```

RESULT 44
HS874C20/c      97847 bp DNA linear PRI 24-JAN-2002
LOCUS      Human DNA sequence from clone RP5-874C20 on chromosome 6p22.1-22.3.
DEFINITION
ACCESSION      HS874C20
VERSION      AL021997.1 GI:3169112
KEYWORDS      HNG; Cpg island; SRE-ZBP; ZFP47; zinc finger protein; ZNF306.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 97847)
AUTHORS      Williams,S.
JOURNAL      Direct Submission
Submitted (04-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On May 30, 1998 this sequence version replaced gi:2916865.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WormBase; Information on the WormBase
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormbase This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP5-874C20 is from the library RPO1-5 constructed by the group of
Pieter de Jong. For further details see

```

| | | |
|---------------|--------------|---|
| repeat_region | 5728..5885 | /note="FRAM repeat: matches 0..154 of consensus" |
| repeat_region | 5887..6015 | /note="MER2 repeat: matches 1..130 of consensus" |
| repeat_region | 6321..6640 | /note="L2 repeat: matches 1156..1523 of consensus" |
| repeat_region | 6644..6827 | /note="MIR repeat: matches 10..234 of consensus" |
| repeat_region | 6828..7152 | /note="LTR18A repeat: matches 1..351 of consensus" |
| repeat_region | 7153..7457 | /note="LTR18A repeat: matches 1..351 of consensus" |
| repeat_region | 7440..7717 | /note="HEVVL18 repeat: matches 1..304 of consensus" |
| repeat_region | 7716..7807 | /note="HEVVL18 repeat: matches 492..770 of consensus" |
| repeat_region | 7795..8156 | /note="HEVVL18 repeat: matches 871..962 of consensus" |
| repeat_region | 8156..8453 | /note="HEVVL18 repeat: matches 871..962 of consensus" |
| repeat_region | 8428..9285 | /note="HEVVL18 repeat: matches 2119..2477 of consensus" |
| repeat_region | 9286..9377 | /note="HEVVL18 repeat: matches 2718..3581 of consensus" |
| repeat_region | 9682..10212 | /note="LTR18B repeat: matches 4394..4486 of consensus" |
| repeat_region | 10268..10420 | /note="LTR18B repeat: matches 1..603 of consensus" |
| repeat_region | 10811..11123 | /note="L2 repeat: matches 1553..1700 of consensus" |
| repeat_region | 11446..11530 | /note="ALUSC repeat: matches 1..307 of consensus" |
| repeat_region | 11531..11752 | /note="MER5B repeat: matches 5..62 of consensus" |
| repeat_region | 11753..11852 | /note="ALUSC repeat: matches 3..228 of consensus" |
| repeat_region | 11856..12304 | /note="MER5B repeat: matches 62..171 of consensus" |
| repeat_region | 12311..12466 | /note="LIMC5 repeat: matches 7375..7793 of consensus" |
| repeat_region | 12315..12480 | /note="L26 copies 6 mer gagaga 59% conserved" |
| repeat_region | 12489..12611 | /note="83 copies 2 mer ga 59% conserved" |
| repeat_region | 12909..13039 | /note="LIMC5 repeat: matches 7810..7929 of consensus" |
| repeat_region | 13461..13656 | /note="LIMD3 repeat: matches 6449..6591 of consensus" |
| repeat_region | 13961..14008 | /note="MIR repeat: matches 12..210 of consensus" |
| repeat_region | 14753..14926 | /note="L24 copies 2 mer ca 83% conserved" |
| repeat_region | 14787..15005 | /note="match: GSS: Em:A0513873" |
| repeat_region | 14787..15005 | /note="match: GSS: Em:A0542521" |
| repeat_region | 14787..14983 | /note="match: GSS: Em:A0542442" |
| repeat_region | 14811..15005 | /note="match: GSS: Em:A0366952" |
| repeat_region | 14812..14994 | /note="match: GSS: Em:A0541753" |
| repeat_region | 15045..15045 | /note="match: GSS: Em:A0541753" |
| repeat_region | 14911..14911 | /note="match: GSS: Em:A082589" |

```

misc_feature      complement(14814..14936)
                  /note="match: GSS: Em:AQ070409"
misc_feature      14817..15021
                  /note="match: GSS: Em:AQ0543269"
misc_feature      /note="match: GSS: Em:AQ0543269"
misc_feature      complement(14823..15015)
                  /note="match: GSS: Em:AQ0729719"
misc_feature      complement(14833..15045)
                  /note="match: GSS: Em:AQ030792"
misc_feature      14833..14937
                  /note="match: GSS: Em:HS203XG5"

Query Match      1.18; Score 23; DB 9; Length 97847;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATTAAGAAAAA 2032
Db 5185 CAATTAAGAAAAA 5163

RESULT 45
AC017637 109151 bp DNA linear HTG 10-DEC-1999
LOCUS      Drosophila melanogaster, *** SEQUENCING IN PROGRESS *** in ordered
DEFINITION pieces.
ACCESSION AC017637
VERSION AC017637.1 GI:6554360
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 109151)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

COMMENT This sequence was identified as CDM:10211565 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a "working draft" sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Source 1..109151
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 31168 a 24117 c 24199 g 29667 t

ORIGIN
Query Match      1.18; Score 23; DB 2; Length 109151;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATTAAGAAAAA 2032
Db 88267 CAATTAAGAAAAA 88289

RESULT 46
AC128720 111669 bp DNA linear HTG 23-JUL-2002
LOCUS      Rattus norvegicus clone CH230-111D9, *** SEQUENCING IN PROGRESS
DEFINITION *** 61 unordered pieces.
ACCESSION AC128720
VERSION AC128720.1 GI:21930183
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE
AUTHORS 1 (bases 1 to 111669)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrooks,S.L., Amaratunga,H.C., Are,J.R., Ayelle,M., Banks,T.,
Barbarta,J., Benton,J., Birnagel,K., Blankenburg,K., Bonini,D.,
Bouck,J., Bowle,S., Burke,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrill,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathore,S.R., David,R.,
David,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagge,N., Ford,J., Foster,P., Frintz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hoogson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,I.F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoen,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtenberg,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,K.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G.,
Oragunye,N., Oyedro,R., Pace,A., Payton,B., Peary,J., Perez,L.,
Peters,L., Pickens,R., Prins,E., Pritchard,M., Ratz,S., Saverly,G.,
Rives,M., Rojas,A., Rojoudokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scheerer,S., Scott,G., Shen,H., Shooshbari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,J., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Teiford,B., Thomas,N., Thomas,S.,
Tsumaki,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

DIRECT SUBMISSION
Unpublished
2 (bases 1 to 111669)
Worley,K.C.

JOURNAL
TITLE Direct Submission
REFERENCE Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
AUTHORS of Molecular and Human Genetics, Baylor College of Medicine, One
JOURNAL Baylor Plaza, Houston, TX 77030, USA

COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GSOI
Center clone name: CH230-111D9
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 45181 bases at least Q40
Consensus quality: 49519 bases at least Q30
Consensus quality: 52605 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/gendank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

* as soon as it is available and the accession number will
* be preserved.
1 1022: contig of 1022 bp in length
1023 1122: gap of unknown length
1123 2588: contig of 1466 bp in length
2589 2688: gap of unknown length
2689 4308: contig of 1620 bp in length
4309 4408: gap of unknown length
4409 5928: contig of 1520 bp in length
5929 6028: gap of unknown length
6029 7515: contig of 1487 bp in length
7516 7615: gap of unknown length
7616 8620: contig of 1005 bp in length
8621 8720: gap of unknown length
8721 9778: contig of 1058 bp in length
9779 9878: gap of unknown length
9879 11361: contig of 1483 bp in length
11362 11461: gap of unknown length
11462 13046: contig of 1585 bp in length
13047 13146: gap of unknown length
13147 14209: contig of 1063 bp in length
14210 14309: gap of unknown length
14310 15521: contig of 1212 bp in length
15522 15621: gap of unknown length
15622 17357: contig of 1736 bp in length
17358 17457: gap of unknown length
17459 18692: contig of 1235 bp in length
18693 18792: gap of unknown length
18793 19964: contig of 1172 bp in length
19965 20064: gap of unknown length
20065 21077: contig of 1013 bp in length
21078 21177: gap of unknown length
21179 22805: contig of 1628 bp in length
22806 24065: gap of unknown length
24066 24165: gap of unknown length
24166 25299: contig of 1134 bp in length
25300 25399: gap of unknown length
25400 26626: contig of 1227 bp in length
26627 26726: gap of unknown length
26727 27816: contig of 1090 bp in length
27817 27916: gap of unknown length
27917 29430: contig of 1514 bp in length
29431 29530: gap of unknown length
29531 31224: contig of 1694 bp in length
31225 31324: gap of unknown length
31325 33253: contig of 1929 bp in length
33254 33353: gap of unknown length
33354 35155: contig of 1802 bp in length
35156 35255: gap of unknown length
35256 36441: contig of 1186 bp in length
36442 36541: gap of unknown length
36542 37601: contig of 1060 bp in length
37602 39355: contig of 1654 bp in length
39356 39455: gap of unknown length
39456 40916: contig of 1461 bp in length
40917 41016: gap of unknown length
41018 42467: contig of 1451 bp in length
42468 42567: gap of unknown length
42569 44379: contig of 1812 bp in length
44380 44479: gap of unknown length
44480 45520: contig of 1041 bp in length
45521 45620: gap of unknown length
45621 47184: contig of 1574 bp in length
47185 47284: gap of unknown length
47285 49498: contig of 2204 bp in length
49499 49598: gap of unknown length
49599 51450: contig of 1852 bp in length
51451 51550: gap of unknown length
51551 53556: contig of 2006 bp in length
53557 53656: gap of unknown length
53657 55732: contig of 2076 bp in length

* 55832: gap of unknown length
55833 58094: contig of 2262 bp in length
58095 58194: gap of unknown length
58195 59938: contig of 1744 bp in length
59939 60038: gap of unknown length
60039 61474: contig of 1436 bp in length
61475 61574: gap of unknown length
61575 63120: contig of 1546 bp in length
63121 63220: gap of unknown length
63221 64264: contig of 1044 bp in length
64265 64364: gap of unknown length
64365 66379: contig of 2015 bp in length
66380 66479: gap of unknown length
66480 67754: contig of 1274 bp in length
67755 67854: gap of unknown length
67855 69537: contig of 1683 bp in length
69538 69636: gap of unknown length
69637 71678: contig of 2042 bp in length
71679 71778: gap of unknown length
71779 73098: contig of 1320 bp in length
73099 73198: gap of unknown length
73199 75335: contig of 2137 bp in length
75336 75435: gap of unknown length
75436 76737: contig of 1302 bp in length
76738 76837: gap of unknown length
76839 78708: contig of 1771 bp in length
78709 81516: contig of 2808 bp in length
81517 81616: gap of unknown length
81617 83645: contig of 2029 bp in length
83646 83745: gap of unknown length
83746 85904: contig of 2159 bp in length
85905 86004: gap of unknown length
86005 88523: contig of 2519 bp in length
88524 88623: gap of unknown length
88624 90881: contig of 2257 bp in length
90882 90980: gap of unknown length
90981 93536: contig of 2556 bp in length
93537 93636: gap of unknown length
93637 95861: contig of 2225 bp in length
95862 95961: gap of unknown length
95962 97882: contig of 1921 bp in length
97883 97982: gap of unknown length

Query Match 1.1%; Score 23; DB 2; Length 111669;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 591 CCTGTGACATCATCCACACAGAT 613
Db 65317 CCTGTGACATCATCCACACAGAT 65339

RESULT 47
AC095664
LOCUS
DEFINITION Rattus norvegicus clone CH230-8L17, *** SEQUENCING IN PROGRESS ***
AC095664
AC095664
VERSION AC095664.3 GI:21722631
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 133902)
AUTHORS Muzny,D.M., Adams,C., Adio-Ogunla,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Boyle,S., Bileva,M., Brown,M., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Fails, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoque, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louleaged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, N., Mapua, P., Martin, R., Martindale, A., Martinez, E., Messer, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabadi, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S., Oguh, M., Okwunye, G., Ocagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojudoan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I., Sodergren, E., Sonakke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umanu, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

 Direct Submission
 2 (bases 1 to 133902)
 Worley, K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 133902)
 Worley, K.C.
 Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:17942230.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc.help@bcm.tmc.edu
 Project Information
 Center project name: GCYV
 Center clone name: CH230-8L17
 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye 100% of reads
 Assembly program: Phrap, version 0.990329
 Consensus quality: 7885 bases at least Q40
 Consensus quality: 86795 bases at least Q30
 Consensus quality: 92571 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 66 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence

* as soon as it is available and the accession number will
 * be preserved.

| | | |
|-------|-------|-----------------------------|
| 1 | 1077 | contig of 1077 bp in length |
| 1078 | 1177 | gap of unknown length |
| 1178 | 2277 | contig of 1100 bp in length |
| 2278 | 2378 | gap of unknown length |
| 2378 | 3709 | contig of 1331 bp in length |
| 3709 | 3808 | gap of unknown length |
| 3809 | 4889 | contig of 1081 bp in length |
| 4890 | 4989 | gap of unknown length |
| 4990 | 5663 | contig of 1674 bp in length |
| 5664 | 6764 | gap of unknown length |
| 6764 | 8020 | contig of 1256 bp in length |
| 8020 | 8119 | gap of unknown length |
| 8120 | 9702 | contig of 1553 bp in length |
| 9703 | 9802 | gap of unknown length |
| 9803 | 11676 | contig of 1874 bp in length |
| 11677 | 11776 | gap of unknown length |
| 11777 | 13246 | contig of 1470 bp in length |
| 13247 | 13347 | gap of unknown length |
| 13347 | 14543 | contig of 1197 bp in length |
| 14544 | 14643 | gap of unknown length |
| 14644 | 15890 | contig of 1247 bp in length |
| 15891 | 15990 | gap of unknown length |
| 15991 | 17240 | contig of 1250 bp in length |
| 17241 | 17340 | gap of unknown length |
| 17341 | 19252 | contig of 1912 bp in length |
| 19253 | 19352 | gap of unknown length |
| 19353 | 20819 | contig of 1457 bp in length |
| 20820 | 20919 | gap of unknown length |
| 20920 | 23340 | contig of 2421 bp in length |
| 23341 | 23440 | gap of unknown length |
| 23441 | 24900 | contig of 1460 bp in length |
| 24901 | 25000 | gap of unknown length |
| 25001 | 26021 | contig of 1021 bp in length |
| 26022 | 26121 | gap of unknown length |
| 26122 | 27281 | contig of 1160 bp in length |
| 27282 | 27381 | gap of unknown length |
| 27382 | 28383 | contig of 1002 bp in length |
| 28384 | 28483 | gap of unknown length |
| 28484 | 30162 | contig of 1679 bp in length |
| 30163 | 30262 | gap of unknown length |
| 30263 | 32073 | contig of 1811 bp in length |
| 32074 | 32173 | gap of unknown length |
| 32174 | 33657 | contig of 1484 bp in length |
| 33658 | 33757 | gap of unknown length |
| 33758 | 36013 | contig of 2256 bp in length |
| 36014 | 36113 | gap of unknown length |
| 36114 | 37587 | contig of 1474 bp in length |
| 37588 | 37687 | gap of unknown length |
| 37688 | 38690 | contig of 1003 bp in length |
| 38691 | 38790 | gap of unknown length |
| 38791 | 40372 | contig of 1582 bp in length |
| 40373 | 40472 | gap of unknown length |
| 40473 | 41611 | contig of 1139 bp in length |
| 41612 | 41711 | gap of unknown length |
| 41712 | 42841 | contig of 1130 bp in length |
| 42842 | 42941 | gap of unknown length |
| 42942 | 44303 | contig of 1362 bp in length |
| 44304 | 44403 | gap of unknown length |
| 44404 | 46227 | contig of 1824 bp in length |
| 46228 | 46327 | gap of unknown length |
| 46328 | 47528 | contig of 1201 bp in length |
| 47529 | 47628 | gap of unknown length |
| 47629 | 49306 | contig of 1678 bp in length |
| 49307 | 49406 | gap of unknown length |
| 49407 | 50779 | contig of 1373 bp in length |
| 50780 | 50879 | gap of unknown length |
| 50880 | 52168 | contig of 1269 bp in length |
| 52169 | 52268 | gap of unknown length |
| 52269 | 53781 | contig of 1513 bp in length |
| 53782 | 53881 | gap of unknown length |
| 53882 | 55307 | contig of 1426 bp in length |

* 55308 55407: gap of unknown length
 * 55408 56741: contig of 1334 bp in length
 * 56742 56841: gap of unknown length
 * 56842 58327: contig of 1486 bp in length
 * 58328 58427: gap of unknown length
 * 58428 60589: contig of 2162 bp in length
 * 60590 60689: gap of unknown length
 * 60690 62255: contig of 1566 bp in length
 * 62256 62355: gap of unknown length
 * 62356 64065: contig of 1710 bp in length
 * 64066 65421: contig of 1256 bp in length
 * 65422 67166: contig of 1645 bp in length
 * 67167 67268: gap of unknown length
 * 67269 68899: contig of 1633 bp in length
 * 68900 71156: contig of 2157 bp in length
 * 71157 71256: gap of unknown length
 * 71257 73078: contig of 1822 bp in length
 * 73079 73178: gap of unknown length
 * 73179 74887: contig of 1709 bp in length
 * 74888 74987: gap of unknown length
 * 74988 77893: contig of 2906 bp in length
 * 77894 77993: gap of unknown length
 * 77994 79712: contig of 1718 bp in length
 * 79712 82367: gap of unknown length
 * 82368 82467: gap of unknown length
 * 82468 84455: contig of 1988 bp in length
 * 84456 85555: gap of unknown length
 * 85556 85795: contig of 1240 bp in length
 * 85796 85895: gap of unknown length
 * 85896 87801: contig of 1906 bp in length
 * 87802 87901: gap of unknown length
 * 87902 89579: contig of 1678 bp in length

Query Match 1.1%; Score 23; DB 2; Length 133902;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 704 GTGCTCAAGAGGTGGCTTCTT 726
 Db 87153 GTGCTCAAGAGGTGGCTTCTT 87175

RESULT 48
 AC116962 141017 bp DNA linear HTG 04-APR-2002
 LOCUS Dictyostelium discoideum chromosome 2 map 1775121-1916136 strain
 DEFINITION A44, *** SEQUENCING IN PROGRESS ***; in ordered pieces.
 AC116962
 VERSION AC116962.1 GI:19920061
 KEYWORDS HTG: HTGS_PHASE2.
 SOURCE Dictyostelium discoideum.
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 1 (bases 1 to 141017)
 Lehnman, R., Baumgart, C., Parra, G., April, J. F., Guiso, R., Kumpf, K.,
 Jungel, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
 Noegel, A. A.
 Sequence and Analysis of Chromosome 2 of Dictyostelium
 Unpublished
 The Dictyostelium Genome Sequencing Consortium
 2 (bases 1 to 141017)
 Baumgart, C.
 Direct Submission
 Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
 CDS predictions from Genaid may contain errors. Further information
 is available from IMB Jena, Department of Genome Analysis
 (http://genome.imb-jena.de/dictyostelium/)

and the University Cologne, Institute for Biochemistry I
 (http://www.uni-koeln.de/dictyostelium/project.shtml
 Funding
 Agency : Deutsche Forschungsgemeinschaft (DFG).
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1. 141017
 /organism="Dictyostelium discoideum"
 /strain="A44"
 /db_xref="taxon:44689"
 /chromosome="2"
 /map="1775121-1916136"
 BASE COUNT 55240 a 15334 c 15477 g 54866 t 100 others
 ORIGIN

Query Match 1.1%; Score 23; DB 2; Length 141017;
 Best Local Similarity 100.0%; Pred. No. 0.48;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032
 Db 28316 CAATATAGAAAAA 28338

RESULT 49
 AC099262/c 143351 bp DNA linear HTG 13-JUL-2002
 LOCUS Rattus norvegicus clone CH230-107H13, *** SEQUENCING IN PROGRESS
 DEFINITION *** 82 unordered pieces.
 AC099262
 VERSION AC099262.4 GI:21731147
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 143351)
 Muzny, D. M., Adams, C., Adio-Oduola, B., Alt-Osman, F. R., Allen, C.,
 Alsbrooks, S. L., Amaral-Tunget, H. C., Are, J. R., Ayala, M., Banks, T.,
 Barbarella, J., Benton, J., Bimoge, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P.,
 Bunay, C., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C.,
 Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C. D., Cox, C., Coyle, M. D., Dettonne, S. R., David, R.,
 Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A.,
 Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H.,
 Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J.,
 Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escoto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J. H., Guevara, R., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B.,
 Housli, F., Howard, S., Huber, J., Huijck, S., Hume, J., Jackson, L. E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvay, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lonsleged, H.,
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maneshari, M., Mapue, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mel, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenko, S., Ogun, M., Okunnu, G.,
 Orangunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojubo, K., Rolfe, M., Ruliz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,

Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, C.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Unpublished
2 (bases 1 to 143351)
Worley, K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 143351)
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17974677.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJXX
Center clone name: CH230-107H13
----- Summary Statistics
Sequencing vector: pLasmid,
Chemistry: Dye-terminator Big Dye 1008 of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 77977 bases at least Q40
Consensus quality: 86348 bases at least Q20
Consensus quality: 92370 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 82 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1041: contig of 1041 bp in length
1141: gap of unknown length
1142 1141: contig of 1071 bp in length
2213 2312: gap of unknown length
2313 3341: contig of 1029 bp in length
3342 3441: gap of unknown length
3442 4442: contig of 1001 bp in length
4443 4542: gap of unknown length
4543 5662: contig of 1120 bp in length
5663 5762: gap of unknown length
5763 7844: contig of 2082 bp in length
7845 7944: gap of unknown length
7945 9672: contig of 1728 bp in length
9673 9772: gap of unknown length
9773 10810: contig of 1038 bp in length
10811 10910: gap of unknown length
10911 11922: contig of 1012 bp in length
11923 12022: gap of unknown length
12023 13037: contig of 1015 bp in length
13038 13137: gap of unknown length
13138 14230: contig of 1093 bp in length
14231 14331: gap of unknown length
14331 15475: contig of 1145 bp in length
15476 15575: gap of unknown length

15576 16609: contig of 1034 bp in length
16610 16709: gap of unknown length
16710 18437: contig of 1728 bp in length
18438 18537: gap of unknown length
18538 19554: contig of 1027 bp in length
19554 19664: gap of unknown length
19665 20756: contig of 1092 bp in length
20757 20856: gap of unknown length
20857 22150: contig of 1294 bp in length
22151 22250: gap of unknown length
22251 23332: contig of 1082 bp in length
23333 23432: gap of unknown length
23433 24747: contig of 1315 bp in length
24748 24847: gap of unknown length
24848 25870: contig of 1022 bp in length
25871 25970: gap of unknown length
25971 27261: contig of 1291 bp in length
27262 27361: gap of unknown length
27362 28490: contig of 1129 bp in length
28491 28590: gap of unknown length
28591 30486: contig of 1896 bp in length
30487 30586: gap of unknown length
30587 32080: contig of 1494 bp in length
32081 32180: gap of unknown length
32181 33385: contig of 1205 bp in length
33386 33485: gap of unknown length
33486 34600: contig of 1115 bp in length
34601 34700: gap of unknown length
34701 36593: contig of 1893 bp in length
36594 37623: gap of unknown length
37623 37863: contig of 1069 bp in length
37863 39209: gap of unknown length
39210 39309: contig of 1347 bp in length
39309 40939: gap of unknown length
40939 41039: contig of 1630 bp in length
41040 42390: gap of unknown length
42390 42490: contig of 1351 bp in length
42491 43515: gap of unknown length
43515 43615: contig of 1025 bp in length
43616 44805: gap of unknown length
44805 44905: contig of 1190 bp in length
44905 44906: gap of unknown length
44906 46765: contig of 1861 bp in length
46765 46866: gap of unknown length
46866 48277: contig of 1411 bp in length
48277 48377: gap of unknown length
48377 49729: contig of 1352 bp in length
49729 49829: gap of unknown length
49829 51737: contig of 1908 bp in length
51737 51837: gap of unknown length
51837 53416: contig of 1579 bp in length
53416 53516: gap of unknown length
53517 55685: contig of 2165 bp in length
55685 55785: gap of unknown length
55786 57414: contig of 1629 bp in length
57415 57515: gap of unknown length
57515 58773: contig of 1255 bp in length
58774 58873: gap of unknown length
58874 60454: contig of 1581 bp in length
60455 60554: gap of unknown length
60555 61862: contig of 1308 bp in length
61863 61962: gap of unknown length
61963 63273: contig of 1311 bp in length
63274 63373: gap of unknown length
63374 65180: contig of 1807 bp in length
65181 65280: gap of unknown length
65281 66544: contig of 1264 bp in length
66545 66644: gap of unknown length
66645 67954: contig of 1310 bp in length
67955 68054: gap of unknown length
68055 69175: contig of 1121 bp in length
69176 69275: gap of unknown length
69276 71084: contig of 1809 bp in length

```

* 71085 71184: gap of unknown length
* 72535 72535: contig of 1351 bp in length
* 72536 72635: gap of unknown length
* 72636 73833: contig of 1198 bp in length
* 73834 73933: gap of unknown length
* 73934 75767: contig of 1834 bp in length
* 75768 77808: gap of unknown length
* 77809 77908: contig of 1941 bp in length
* 77909 79721: contig of 1813 bp in length

Query Match      1.1%: Score 23; DB 2: Length 143351;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 572 GCCCTGCTCTGCACCTGCCTG 594
Db 132644 GCCCTGCTCTGCACCTGCCTG 132622

RESULT 50
AC115914 159170 bp DNA linear HTG 20-JUN-2002
LOCUS AC115914
DEFINITION Mus musculus clone RP24-487J5, WORKING DRAFT SEQUENCE, 16 ordered
pieces.
AC115914.2 GI:21490457
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-487J5
JOURNAL Unpublished
2 (bases 1 to 159170)
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gilde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menues,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhthang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strausman,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trifillio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 159170)
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhaltier,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gilde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczeky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menues,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhthang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strausman,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trifillio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 159170)

```

```

TITLE
JOURNAL
COMMENT
Liu,G., Maclean,C., MacDonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,
Menues,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunhthang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strausman,N., Subramanian,A., Talamas,J., Testaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trifillio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL Submitted (20-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 20, 2002 this sequence version replaced g1.119683593.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 124872
Center clone name: 487.J.5
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 154428 bases at least Q40
Consensus quality: 156177 bases at least Q30
Consensus quality: 156924 bases at least Q20
Insert size: 163000; agarose-ff
Insert size: 157670; sum-of-contigs
Quality coverage: 6.4 in Q20 bases; sum-of-contigs
Quality coverage: 6.6 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 16 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
1 1032: contig of 1032 bp in length
* 1033 1132: gap of 100 bp
* 1133 4008: contig of 2876 bp in length
* 4009 4108: gap of 100 bp
* 4109 6871: contig of 2763 bp in length
* 6872 6971: gap of 100 bp
* 6972 8935: contig of 1964 bp in length
* 8936 9035: gap of 100 bp
* 9036 13009: contig of 3974 bp in length
* 13010 13109: gap of 100 bp
* 13110 15598: contig of 2489 bp in length
* 15599 15698: gap of 100 bp
* 15699 19053: contig of 3355 bp in length
* 19054 19153: gap of 100 bp
* 19154 26269: contig of 7116 bp in length
* 26270 26369: gap of 100 bp
* 26370 34645: contig of 8276 bp in length
* 34646 34745: gap of 100 bp
* 34746 46455: contig of 11710 bp in length
* 46456 46555: gap of 100 bp
* 46556 55127: contig of 8872 bp in length
* 55128 55227: gap of 100 bp
* 55228 67686: contig of 12455 bp in length
* 67687 67786: gap of 100 bp
* 67787 90042: contig of 22256 bp in length

```

* 9043 90142: gap of 100 bp
* 90143 109818: contig of 19676 bp in length
* 109819 109918: gap of 100 bp
* 109919 130304: contig of 20386 bp in length
* 130305 130404: gap of 100 bp
* 130405 159170: contig of 28766 bp in length.
Location/Qualifiers

FEATURES
source

1. 159170
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_id="RP24-487J5"
/clone_lib="RP24 Male Mouse BAC"
1. 1032
/note="assembly_fragment"
113. 4008
/note="assembly_fragment"
4109. 6871
/note="assembly_fragment"
6972. 8935
/note="assembly_fragment"
9036. 13009
/note="assembly_fragment"
13110. 15598
/note="assembly_fragment"
15699. 18053
/note="assembly_fragment"
19154. 26269
/note="assembly_fragment"
26370. 34645
/note="assembly_fragment"
34746. 46455
/note="assembly_fragment"
48556. 55127
/note="assembly_fragment"
55228. 67686
/note="assembly_fragment"
67787. 90042
/note="assembly_fragment"
90143. 109818
/note="assembly_fragment"
109919. 130304
/note="assembly_fragment"
130405. 159170
/note="assembly_fragment"
BASE COUNT 41889 a 35541 c 36173 g 44061 t 1506 others
ORIGIN

Query Match 1.16; Score 23; DB 2; Length 159170;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 704 GTGCTCAGAGAGTCCGCTTCTT 726
|||||
Db 145557 GTGCTCAGAGAGTCCGCTTCTT 145579

Search completed: January 11, 2003, 02:09:57
Job time : 11357 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:56:19 ; Search time 427 Seconds
(without alignments)
10716.770 Million cell updates/sec

Title: US-09-816-825-1
Perfect score: 2032
Sequence: 1 ggcctcgcagccagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 20

Total number of hits satisfying chosen parameters: 145

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-----------------------------|
| 1 | 2032 | 100.0 | 2032 | 20 | AAZ20792 Human glycosyl sul |
| 2 | 1802 | 88.7 | 1979 | 22 | AAK94229 Human full-length |
| 3 | 1615 | 79.5 | 2065 | 21 | AAZ94211 Human transferrase |
| 4 | 1211 | 59.6 | 1333 | 24 | AA516947 Human L-selectin s |
| 5 | 741 | 36.5 | 877 | 22 | AAK91803 Human CDNA 5'-end |
| 6 | 741 | 36.5 | 877 | 22 | AAK93921 Human CDNA clone r |
| 7 | 505 | 24.9 | 517 | 22 | AAK54724 Human colon cancer |
| 8 | 351 | 17.3 | 548 | 22 | AAK92588 Human CDNA 3'-end |
| 9 | 306 | 15.1 | 2988 | 21 | AAK61556 Human ORFX ORP1711 |

| | | | | | |
|----|-----|------|--------|----|------------------------------|
| 10 | 286 | 14.1 | 389 | 24 | AAK54794 Human colon cancer |
| 11 | 60 | 3.0 | 60 | 24 | AAK38074 Human spliced tran |
| 12 | 59 | 2.9 | 1647 | 24 | AAZ24670 Human drug metabol |
| 13 | 59 | 2.9 | 1694 | 22 | AAZ02700 Human glycosyl sul |
| 14 | 59 | 2.9 | 2044 | 22 | AAZ02699 Human glycosyl sul |
| 15 | 59 | 2.9 | 2170 | 22 | AAZ02898 Human glycosyl sul |
| 16 | 59 | 2.9 | 2544 | 24 | AAK89506 Human corneal N-ac |
| 17 | 59 | 2.9 | 48436 | 24 | AAK89506 Human corneal N-ac |
| 18 | 59 | 2.9 | 160552 | 22 | AAZ02697 Human glycosyl sul |
| 19 | 37 | 1.8 | 30 | 24 | AAZ20798 PCR primer for gly |
| 20 | 37 | 1.5 | 37 | 24 | AAZ20798 PCR primer for gly |
| 21 | 27 | 1.3 | 1926 | 20 | AAZ20793 Mouse glycosyl sul |
| 22 | 23 | 1.1 | 616 | 21 | AAK74310 Human secreted pro |
| 23 | 23 | 1.1 | 1142 | 22 | AAK41008 cDNA encoding nove |
| 24 | 23 | 1.1 | 1937 | 24 | AAK16948 Murine intestinal- |
| 25 | 23 | 1.1 | 1989 | 22 | AAZ02696 Mouse glycosyl sul |
| 26 | 23 | 1.1 | 2411 | 23 | AAV29163 Human prostate exp |
| 27 | 23 | 1.1 | 2860 | 18 | AAK63303 A. chrysogenum O-a |
| 28 | 23 | 1.1 | 4871 | 23 | AAK16362 Drosophila melanog |
| 29 | 23 | 1.1 | 5101 | 23 | AAK16364 Drosophila melanog |
| 30 | 23 | 1.1 | 10138 | 23 | AAK16360 Drosophila melanog |
| 31 | 22 | 1.1 | 22 | 24 | AAK70834 PCR primer for hum |
| 32 | 22 | 1.1 | 22 | 24 | AAK16960 Human L-selectin s |
| 33 | 22 | 1.1 | 22 | 24 | AAK16962 Human L-selectin s |
| 34 | 22 | 1.1 | 22 | 24 | AAK16964 Human L-selectin s |
| 35 | 22 | 1.1 | 341 | 23 | AAK36642 Human prostate exp |
| 36 | 22 | 1.1 | 502 | 23 | AAK34232 Human prostate exp |
| 37 | 22 | 1.1 | 1781 | 22 | AAK41556 cDNA encoding nove |
| 38 | 22 | 1.1 | 1782 | 24 | AAK63555 Rat sequence diffe |
| 39 | 22 | 1.1 | 2415 | 22 | AAH75633 Human ribosomal pr |
| 40 | 22 | 1.1 | 3338 | 12 | AAO11712 Shuttle vector PMU |
| 41 | 22 | 1.1 | 3597 | 22 | AAK82684 HS0800936 gene SEQ |
| 42 | 22 | 1.1 | 73947 | 23 | AAK07230 Drosophila melanog |
| 43 | 21 | 1.0 | 21 | 24 | AAK70835 PCR primer for hum |
| 44 | 21 | 1.0 | 21 | 24 | AAK16961 Human L-selectin s |
| 45 | 21 | 1.0 | 51 | 22 | AAK27808 Human SNP oligonuc |
| 46 | 21 | 1.0 | 262 | 23 | AAK08745 Human prostate exp |
| 47 | 21 | 1.0 | 294 | 22 | AAK63356 DNA encoding larva |
| 48 | 21 | 1.0 | 397 | 23 | AAK56821 Human prostate exp |
| 49 | 21 | 1.0 | 405 | 22 | AAK92205 Human polynucleoti |
| 50 | 21 | 1.0 | 411 | 24 | AAK94940 Gene #1438 used to |
| 51 | 21 | 1.0 | 411 | 24 | AAK62995 Breast cancer rela |
| 52 | 21 | 1.0 | 428 | 23 | AAK46141 Human prostate exp |
| 53 | 21 | 1.0 | 458 | 23 | AAK83335 DNA encoding novel |
| 54 | 21 | 1.0 | 733 | 22 | AAK19661 Human neuroblastom |
| 55 | 21 | 1.0 | 785 | 24 | AAK96921 Arabidopsis thalia |
| 56 | 21 | 1.0 | 986 | 24 | AAK98495 Human cancer assoc |
| 57 | 21 | 1.0 | 998 | 21 | AAK77903 Human colon specif |
| 58 | 21 | 1.0 | 1563 | 24 | AAK073783 Human secreted pro |
| 59 | 21 | 1.0 | 1461 | 21 | AAK78402 Human CDNA encodin |
| 60 | 21 | 1.0 | 1469 | 21 | AAK10276 Sequence encoding |
| 61 | 21 | 1.0 | 1493 | 11 | AAK00492 Human osteopontin |
| 62 | 21 | 1.0 | 1493 | 20 | AAK33994 Human cancer assoc |
| 63 | 21 | 1.0 | 1792 | 21 | AAK77865 Human prostate can |
| 64 | 21 | 1.0 | 1792 | 22 | AAK72777 Human cancer assoc |
| 65 | 21 | 1.0 | 2013 | 21 | AAK78178 Human cancer assoc |
| 66 | 21 | 1.0 | 2655 | 22 | AAK72751 Human prostate can |
| 67 | 21 | 1.0 | 2835 | 22 | AAK52037 Human polynucleoti |
| 68 | 21 | 1.0 | 6608 | 24 | AAK34034 Human immune syst |
| 69 | 21 | 1.0 | 10004 | 22 | AAK14483 Human nervous syst |
| 70 | 21 | 1.0 | 10246 | 22 | AAK14485 Human nervous syst |
| 71 | 20 | 1.0 | 20 | 22 | AAK50713 PolyPyrimidine Crt |
| 72 | 20 | 1.0 | 88 | 21 | AAK25980 Human ovarian can |
| 73 | 20 | 1.0 | 161 | 24 | AAK57865 Human prostate exp |
| 74 | 20 | 1.0 | 225 | 23 | AAK25536 Human prostate exp |
| 75 | 20 | 1.0 | 293 | 22 | AAK09382 Rat allograft infl |
| 76 | 20 | 1.0 | 295 | 21 | AAK65602 Rat allograft infl |
| 77 | 20 | 1.0 | 308 | 22 | AAK16693 Human breast cance |
| 78 | 20 | 1.0 | 362 | 16 | AAK09378 Rat allograft infl |
| 79 | 20 | 1.0 | 362 | 21 | AAK65598 Rat allograft infl |
| 80 | 20 | 1.0 | 364 | 22 | AAK07790 Human breast cance |
| 81 | 20 | 1.0 | 384 | 23 | AAK13665 Human prostate exp |
| 82 | 20 | 1.0 | | | |

| | | | | | | |
|-----|----|-----|--------|----|----------|---------------------|
| 83 | 20 | 1.0 | 387 | 20 | AA420849 | Secreted protein E |
| 84 | 20 | 1.0 | 403 | 22 | AA187736 | Human polynucleoti |
| 85 | 20 | 1.0 | 405 | 23 | ABV34779 | Human prostate exp |
| 86 | 20 | 1.0 | 405 | 23 | ABV43630 | Human prostate exp |
| 87 | 20 | 1.0 | 410 | 23 | ABV18937 | Human prostate exp |
| 88 | 20 | 1.0 | 410 | 23 | ABK54824 | Human colon cancer |
| 89 | 20 | 1.0 | 435 | 23 | ABV56080 | Human prostate exp |
| 90 | 20 | 1.0 | 451 | 23 | ABV48716 | Human prostate exp |
| 91 | 20 | 1.0 | 468 | 22 | ABA14197 | Human nervous syst |
| 92 | 20 | 1.0 | 474 | 22 | AA560615 | Human cancer agent |
| 93 | 20 | 1.0 | 488 | 22 | AA529035 | CDNA encoding for |
| 94 | 20 | 1.0 | 517 | 23 | ABK62399 | Rat sequence diffie |
| 95 | 20 | 1.0 | 527 | 23 | ABV58675 | Human prostate exp |
| 96 | 20 | 1.0 | 554 | 24 | ABK54892 | Human colon cancer |
| 97 | 20 | 1.0 | 553 | 22 | AA113827 | Human breast cancer |
| 98 | 20 | 1.0 | 630 | 21 | AA116113 | Human prostate can |
| 99 | 20 | 1.0 | 804 | 22 | AA197306 | Human neuroblastom |
| 100 | 20 | 1.0 | 873 | 22 | AA122693 | Human breast cancer |
| 101 | 20 | 1.0 | 977 | 21 | AA564465 | Human prostate can |
| 102 | 20 | 1.0 | 986 | 24 | ABQ34314 | Oligonucleotide fo |
| 103 | 20 | 1.0 | 986 | 24 | ABQ34315 | Oligonucleotide fo |
| 104 | 20 | 1.0 | 1094 | 24 | AA596308 | Arabidopsis CDNA e |
| 105 | 20 | 1.0 | 1104 | 24 | AA580708 | Soybean chlorophyl |
| 106 | 20 | 1.0 | 1117 | 21 | AA118221 | Lung cancer associ |
| 107 | 20 | 1.0 | 1243 | 21 | AAQ31337 | Arabidopsis thalia |
| 108 | 20 | 1.0 | 1297 | 13 | AAQ21645 | 3' coding sequence |
| 109 | 20 | 1.0 | 1316 | 22 | AA472877 | Human secreted pro |
| 110 | 20 | 1.0 | 1389 | 21 | AA339055 | Human secreted pro |
| 111 | 20 | 1.0 | 1504 | 21 | AA38952 | Arabidopsis thalia |
| 112 | 20 | 1.0 | 1526 | 21 | AACT6491 | Human OREX ORE2046 |
| 113 | 20 | 1.0 | 1587 | 21 | AA34626 | Drosophila dm11 t |
| 114 | 20 | 1.0 | 1708 | 21 | AACT4428 | Human secreted pro |
| 115 | 20 | 1.0 | 1770 | 14 | AAQ37684 | P. falciparum antiq |
| 116 | 20 | 1.0 | 1770 | 14 | AAQ37124 | Human INTRCPT 25 |
| 117 | 20 | 1.0 | 1869 | 22 | AA444878 | Human secreted pro |
| 118 | 20 | 1.0 | 1869 | 22 | AA45014 | Human secreted pro |
| 119 | 20 | 1.0 | 1869 | 22 | AA45015 | Human secreted pro |
| 120 | 20 | 1.0 | 1869 | 22 | AA45016 | Human secreted pro |
| 121 | 20 | 1.0 | 1869 | 22 | AA45017 | Human secreted pro |
| 122 | 20 | 1.0 | 2290 | 22 | AA59490 | Frog CDNA encoding |
| 123 | 20 | 1.0 | 2429 | 22 | AA73413 | Grand fir monoterp |
| 124 | 20 | 1.0 | 2441 | 24 | AA562635 | CDNA sequence #422 |
| 125 | 20 | 1.0 | 2843 | 24 | ABK48708 | CDNA encoding huma |
| 126 | 20 | 1.0 | 2888 | 21 | AA96483 | CDNA encoding a hu |
| 127 | 20 | 1.0 | 2894 | 20 | AA86127 | EST sequence for D |
| 128 | 20 | 1.0 | 2959 | 24 | AA566336 | Arabidopsis CDNA e |
| 129 | 20 | 1.0 | 3492 | 20 | AA53745 | CDNA encoding a pr |
| 130 | 20 | 1.0 | 3515 | 22 | AA75343 | Human TGF-beta rec |
| 131 | 20 | 1.0 | 3763 | 19 | AA58194 | Human myosin I-cha |
| 132 | 20 | 1.0 | 3763 | 23 | AA569872 | DNA encoding novel |
| 133 | 20 | 1.0 | 5814 | 24 | ABN87855 | Human ovary specif |
| 134 | 20 | 1.0 | 7138 | 24 | ABK28455 | DNA transcription |
| 135 | 20 | 1.0 | 8652 | 23 | ABL03026 | Drosophila melanog |
| 136 | 20 | 1.0 | 9516 | 22 | AAK73555 | Human immune/haema |
| 137 | 20 | 1.0 | 9516 | 22 | AAK81394 | Human immune/haema |
| 138 | 20 | 1.0 | 13192 | 22 | AAK73556 | Human immune/haema |
| 139 | 20 | 1.0 | 13192 | 22 | AAK81396 | Human immune/haema |
| 140 | 20 | 1.0 | 14482 | 23 | ABL10086 | Drosophila melanog |
| 141 | 20 | 1.0 | 20990 | 23 | ABL03102 | Drosophila melanog |
| 142 | 20 | 1.0 | 35959 | 22 | AAK78275 | Human immune/haema |
| 143 | 20 | 1.0 | 38653 | 22 | AA544513 | Human LEX1 DNA cl |
| 144 | 20 | 1.0 | 147419 | 22 | ABK83574 | Human CDNA differe |
| 145 | 20 | 1.0 | 349980 | 22 | ABK86431 | Pyrococcus abyssi |

ALIGNMENTS

RESULT 1
AA220792
ID AA220792 standard: DNA: 2032 BP.
XX
AC AA220792:

| | | |
|----------------------------|---|---------------------------|
| XX | 08-DEC-1999 (first entry) | 08-DEC-1999 (first entry) |
| DT | Human glycosyl sulfotransferase-3 coding sequence. | |
| XX | | |
| DE | Human glycosyl sulfotransferase-3 coding sequence. | |
| XX | | |
| KW | Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing; | |
| KM | selectin binding interaction; inflammation; lymphocyte homing; human; | |
| KW | secondary lymph organ; ss. | |
| OS | Homo sapiens. | |
| XX | | |
| PN | W09949018-A1. | |
| PD | 30-SEP-1999. | |
| XX | | |
| PF | 26-FEB-1999; 99WO-US04316. | |
| XX | | |
| PR | 20-MAR-1998; 98US-0045284. | |
| PR | 12-NOV-1998; 98US-0190911. | |
| XX | | |
| PA | (RECC) UNIV CALIFORNIA. | |
| PA | (SYNT) SYNTX USA INC. | |
| PI | Bistrup A, Rosen SD, Tangemann K, Hemmerich S; | |
| DR | WPI: 1999-580442/49. | |
| DR | P-PSDB: AAY39918. | |
| PT | Human and murine glycosyl sulfotransferase 3 and related | |
| PS | polynucleotides | |
| XX | Claim 4; Fig 1; 59pp: English. | |
| CC | This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of | |
| CC | the invention. The nucleic acid sequences, probes and primers derived | |
| CC | from these, proteins and antibodies are useful in detecting homologues. | |
| CC | The sequences, antibodies and methods are useful in the diagnosis and | |
| CC | treatment of diseases associated with selectin binding interactions, | |
| CC | including conditions associated with or resulting from the homing of | |
| CC | leukocytes to sites of inflammation and the normal homing of lymphocytes | |
| CC | to secondary lymph organs. | |
| XX | | |
| SQ | Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other: | |
| Query Match | 100.0%; Score 2032; DB 20; Length 2032; | |
| Best Local Similarity | 100.0%; Pred. No. 0; | |
| Matches 2032; Conservative | 0; Mismatches 0; Indels 0; Gaps 0; | |
| QY | 1 GGCTGAGGCCAGGATGCTCCAGTCTGGGGGAAATGCTTCTATTTGCTTCCAG 60 | |
| DB | 1 GGCTGAGGCCAGGATGCTCCAGTCTGGGGGAAATGCTTCTATTTGCTTCCAG 60 | |
| QY | 61 CCCACCTCAGCAGTCTCCACCCCTTGAGTCTCAGAGTGTAAAGCTGTACTTCA 120 | |
| DB | 61 CCCACCTCAGCAGTCTCCACCCCTTGAGTCTCAGAGTGTAAAGCTGTACTTCA 120 | |
| QY | 121 CAGCTTCTCGGAGGAGTCTTCTCAAGCCGCTTTCAGAGTCTTCCACTTCAGCAC 180 | |
| DB | 121 CAGCTTCTCGGAGGAGTCTTCTCAAGCCGCTTTCAGAGTCTTCCACTTCAGCAC 180 | |
| QY | 181 AATGCTACGCTTAAATAAATGAAGCTCTGTTGTGTTGCCAGAGGCCATCTT 240 | |
| DB | 181 AATGCTACGCTTAAATAAATGAAGCTCTGTTGTGTTGCCAGAGGCCATCTT 240 | |
| QY | 241 GGCTATATCTTCCACATGTACAGCCACATCAGTCCCTGTATATGAAGGACACCC 300 | |
| DB | 241 GGCTATATCTTCCACATGTACAGCCACATCAGTCCCTGTATATGAAGGACACCC 300 | |
| QY | 301 CGAGGCGATGCAAGTCTGTTCTCTTCTTCCGCGCTCTGCTTTTGTGGGCA 360 | |
| DB | 301 CGAGGCGATGCAAGTCTGTTCTCTTCTTCCGCGCTCTGCTTTTGTGGGCA 360 | |
| QY | 361 GCTTTTGGGCGAGCACCCAGATGTTTCTACGTATGAGCCCGGCTGACAGTGTGAT 420 | |


```
Db 361 GCTTTTGGGAGACCCAGATGTTTCTTACCTATGAGAGCCCGCTGGACGCTGTGGAT 420
QY 421 GACCTTCAGAGAGACACCCGCTGATGCTCAGATGAGCTGTGGGAGATGATAGCGGC 480
Db 421 GACCTTCAGAGAGACACCCGCTGATGCTCAGATGAGCTGTGGGAGATGATAGCGGC 480
QY 481 GCTCTTCTTGGGACATGAGCGCTTTGATGCTTACATGAGACCTGTGCCGAGACA 540
Db 481 GCTCTTCTTGGGACATGAGCGCTTTGATGCTTACATGAGACCTGTGCCGAGACA 540
QY 541 GTCCAGCCTCTTTCAGTGGGAGAACAGCGCGGCTGTGTCTCTCAGCTGCTGTGACAT 600
Db 541 GTCCAGCCTCTTTCAGTGGGAGAACAGCGCGGCTGTGTCTCTCAGCTGCTGTGACAT 600
QY 601 CATCCACAGATGAAATCATCCCGGGGCTCACTGCAAGGCTCCTGTGACATCAACAGCC 660
Db 601 CATCCACAGATGAAATCATCCCGGGGCTCACTGCAAGGCTCCTGTGACATCAACAGCC 660
QY 661 CTTTGAGTGGTGGAGAGAGCGCTCCGCTCTTACAGCCAGCTGGTGTCAAGAGGTGGC 720
Db 661 CTTTGAGTGGTGGAGAGAGCGCTCCGCTCTTACAGCCAGCTGGTGTCAAGAGGTGGC 720
QY 721 CTTTCTTCAACCTGCAAGTCCCTCTACCCGCTGCTGAAGACCCCTCCCTCAACCTGATAT 780
Db 721 CTTTCTTCAACCTGCAAGTCCCTCTACCCGCTGCTGAAGACCCCTCCCTCAACCTGATAT 780
QY 781 CGTGACCTGCTCCGGGAGACCCCGGCGCTGTCCGTTCCCGAGACGACACAAAGGAGGA 840
Db 781 CGTGACCTGCTCCGGGAGACCCCGGCGCTGTCCGTTCCCGAGAGCGACAAAGGAGGA 840
QY 841 TCTGATGATGACAGTGGCATTTGATGAGGAGACATGAGAGAACTCAAGAGAGAGA 900
Db 841 TCTGATGATGACAGTGGCATTTGATGAGGAGACATGAGAGAACTCAAGAGAGAGA 900
QY 901 CCAACCCCTACTATGTGATGACAGTCACTGTGCAAGCCAGCTGAGATCTACAAGACAT 960
Db 901 CCAACCCCTACTATGTGATGACAGTCACTGTGCAAGCCAGCTGAGATCTACAAGACAT 960
QY 961 CCAGTCCCTTGCCCAAGGCGCTGCAAGAGACGCTACTGCTGTGGCTATAGAGACCTGGC 1020
Db 961 CCAGTCCCTTGCCCAAGGCGCTGCAAGAGACGCTACTGCTGTGGCTATAGAGACCTGGC 1020
QY 1021 TCGAGCCCTGCTGGCCAGACTGCCGATGATGAATTCGTGGGATGGAATCTTGGC 1080
Db 1021 TCGAGCCCTGCTGGCCAGACTGCCGATGATGAATTCGTGGGATGGAATCTTGGC 1080
QY 1081 CCATCTTCAGACCTGGGTGGCATTAATCACCAGGCAAGGGCATGGGTGACACGCTTT 1140
Db 1081 CCATCTTCAGACCTGGGTGGCATTAATCACCAGGCAAGGGCATGGGTGACACGCTTT 1140
QY 1141 CCACACAATGCCAGGGATGCCCTTAATGTCTCCAGGCTGGCGCTGTGGCCCTA 1200
Db 1141 CCACACAATGCCAGGGATGCCCTTAATGTCTCCAGGCTGGCGCTGTGGCCCTA 1200
QY 1201 TGAAGAGTTCCTGACTTCGAAGAGCCTGTGGCATGCAATTTGCTGGGCTACCG 1260
Db 1201 TGAAGAGTTCCTGACTTCGAAGAGCCTGTGGCATGCAATTTGCTGGGCTACCG 1260
QY 1261 CCAGCTTCGATCTGACACAGAGAGAAACCTGTGCTGATTTCTGTCTACTGAGAC 1320
Db 1261 CCAGCTTCGATCTGACACAGAGAGAAACCTGTGCTGATTTCTGTCTACTGAGAC 1320
QY 1321 TGTCCCTGAGCAAAATCCACTAGAGGGGTGAAGAGGCTTGGTCCACCTGGTGTGACC 1380
Db 1321 TGTCCCTGAGCAAAATCCACTAGAGGGGTGAAGAGGCTTGGTCCACCTGGTGTGACC 1380
QY 1381 TCACTGACTTCTCTGATAGCTTCTGAGCCTGTGCTACATCTGAGACCTTAATACATG 1440
Db 1381 TCACTGACTTCTCTGATAGCTTCTGAGCCTGTGCTACATCTGAGACCTTAATACATG 1440
QY 1441 TCTGTGGGTATCACAGTGTGATGTGTGTCCACAGCTGCTCAACAGAGACTTTT 1500
Db 1441 TCTGTGGGTATCACAGTGTGATGTGTGTGTCCACAGCTGCTCAACAGAGACTTTT 1500
```

```
Db 1441 TCTGTGGGTATCACAGTGTGATGTGTGTGTCCACAGCTGCTCAACAGAGACTTTT 1500
QY 1501 GTGTCCATGCTTGTGTCTAGAAAACAGACTGGGGAACCTTATGTGAGCAGCATCCAC 1560
Db 1501 GTGTCCATGCTTGTGTCTAGAAAACAGACTGGGGAACCTTATGTGAGCAGCATCCAC 1560
QY 1561 CAGTGAACAGGGGTATGCTCTTCTTTCTTTCTTGAATCTTCTGCTGTGGGACACTTCAG 1620
Db 1561 CAGTGAACAGGGGTATGCTCTTCTTTCTTTCTTGAATCTTCTGCTGTGGGACACTTCAG 1620
QY 1621 AGACTTGTGGCTGGAGGCTATTAGACGAGACAGTATCGAGGAAATGATCCATTA 1680
Db 1621 AGACTTGTGGCTGGAGGCTATTAGACGAGACAGTATCGAGGAAATGATCCATTA 1680
QY 1681 ACCTCCCTGTCCACATCTTTCGCCAATGGGGAATGATCTTTTCACAAAGCTCACACAC 1740
Db 1681 ACCTCCCTGTCCACATCTTTCGCCAATGGGGAATGATCTTTTCACAAAGCTCACACAC 1740
QY 1741 ATTTTCACAGAGATGCAATTTCTGAGCCCTTGGAGTTCCCASTGATTCAGAGAGAA 1800
Db 1741 ATTTTCACAGAGATGCAATTTCTGAGCCCTTGGAGTTCCCASTGATTCAGAGAGAA 1800
QY 1801 GTGGAAACAAGTGTGATGCTTACTATGAGCTTACACATCAGAGCTATCGTAATCAGA 1860
Db 1801 GTGGAAACAAGTGTGATGCTTACTATGAGCTTACACATCAGAGCTATCGTAATCAGA 1860
QY 1861 AATATGAACAAAATCTCTGACAAAAGAGCAAGCTTTAAGTTACACAGGGTCTGGGC 1920
Db 1861 AATATGAACAAAATCTCTGACAAAAGAGCAAGCTTTAAGTTACACAGGGTCTGGGC 1920
QY 1921 TGCATTTGAATACACTTCCCTCTGATTTTCCCTACATGAGAGACTTTGACCTGTG 1980
Db 1921 TGCATTTGAATACACTTCCCTCTGATTTTCCCTACATGAGAGACTTTGACCTGTG 1980
QY 1981 AAGCTGCCATCTGTTAATACTAAATTCCTAATAAGAAAAA 2032
Db 1981 AAGCTGCCATCTGTTAATACTAAATTCCTAATAAGAAAAA 2032

RESULT 2
AAK94229
ID AAK94229 standard; cDNA; 1979 BP.
XX
AC AAK94229;
XX
DI 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 2816.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT P-PSDB; AAK93309.
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
```

XX Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.
PS
XX The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.
CC
XX
SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;
Query Match 88.7%; Score 1802; DB 22; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 114 ACTTCAAGCTTCTGAGAGCGAGTCTTCAAGCCGCTTGTGAGAGTCTTCCACT 173
DB 1 ACTTCAAGCTTCTGAGAGCGAGTCTTCAAGCCGCTTGTGAGAGTCTTCCACT 60
QY 174 TCAGACAAATGCTACTGCTTAAAAAATGAGTCTGCTTGTGAGAGTCTTCCAGATG 233
DB 61 TCAGACAAATGCTACTGCTTAAAAAATGAGTCTGCTTGTGAGAGTCTTCCAGATG 120
QY 234 CCATCTGGCTCTATCTTCCACATGTACAGCAACAACATAGCTCCCTGTCTATGAGG 293
DB 121 CCATCTGGCTCTATCTTCCACATGTACAGCAACAACATAGCTCCCTGTCTATGAGG 180
QY 224 CACAGCCGAGCGATGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 353
DB 181 CACAGCCGAGCGATGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 354 TGGGCGACCTTTTGGGCGACGACCCAGATGTTTCTACTGATGAGCCGCTGCGACG 413
DB 241 TGGGCGACCTTTTGGGCGACGACCCAGATGTTTCTACTGATGAGCCGCTGCGACG 300
QY 414 TGTGATGACCTTAACAGACGACCCGCTGATGCTGACATAGGCTGCTGAGATCTGA 473
DB 301 TGTGATGACCTTAACAGACGACCCGCTGATGCTGACATAGGCTGCTGAGATCTGA 360
QY 474 TACGGGCGCTTCTTGTGAGACATGAGCGCTTGTGATGCTTACATAGGAACCTGCTCC 533
DB 361 TACGGGCGCTTCTTGTGAGACATGAGCGCTTGTGATGCTTACATAGGAACCTGCTCC 420
QY 534 GGAGACAGTCAGCTCTTTCAGTGGGAAACAGCCGCGCTGTGTGCTGACCTGCTCT 593
DB 421 GGAGACAGTCAGCTCTTTCAGTGGGAAACAGCCGCGCTGTGTGCTGACCTGCTCT 480
QY 594 GTGACATCATCCCAAGATGAATCATCCCGGCGCTCACTGAGGCTCTGTGAGATC 653
DB 481 GTGACATCATCCCAAGATGAATCATCCCGGCGCTCACTGAGGCTCTGTGAGATC 540
QY 654 AACAGCCCTTGAAGTGTGAGAGAGCGCTGCGCTTCTCAACGACAGTGTGCTCAAG 713
DB 541 AACAGCCCTTGAAGTGTGAGAGAGCGCTGCGCTTCTCAACGACAGTGTGCTCAAG 600
QY 714 AGTGGCCTTCTTCAACCTGCACTCCCTCAACCGCTGCTGGAAGAACCTCCCTCAAC 773
DB 601 AGTGGCCTTCTTCAACCTGCACTCCCTCAACCGCTGCTGGAAGAACCTCCCTCAAC 660
QY 774 TGCATATCTGCACTGCTGCGGAGCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 833
DB 661 TGCATATCTGCACTGCTGCGGAGCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 834 AGGAGATCTCATGATGAGAGTGTGATGAGGCGCATGAGCAAACTCAAGA 893
DB 721 AGGAGATCTCATGATGAGAGTGTGATGAGGCGCATGAGCAAACTCAAGA 780

QY 894 AGAGGACCAACCTTACTATGTGATGACAGTCTTCCCAAGCCAGCTGAGATCTCA 953
DB 781 AGAGGACCAACCTTACTATGTGATGACAGTCTTCCCAAGCCAGCTGAGATCTCA 840
QY 954 AGACATCCAGCTTGTGCGCAAGGCGCTGACAGAGGCTTGTGCGCTTGTGAGG 1013
DB 841 AGACATCCAGCTTGTGCGCAAGGCGCTGACAGAGGCTTGTGCGCTTGTGAGG 900
QY 1014 ACTGCGCTGAGCCCTGTGCGCCAGACTTCCGATGTATGAAATGCTGAGATGAT 1073
DB 901 ACTGCGCTGAGCCCTGTGCGCCAGACTTCCGATGTATGAAATGCTGAGATGAT 960
QY 1074 TCTTGGCCCATTTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1133
DB 961 TCTTGGCCCATTTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1134 AGCTTTTCCACAAATGCGACGAGTGTGCTTATGTCTCCAGGCTTGGCGCTGCT 1193
DB 1021 AGCTTTTCCACAAATGCGACGAGTGTGCTTATGTCTCCAGGCTTGGCGCTGCT 1080
QY 1194 TGGCTATGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1253
DB 1081 TGGCTATGAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 1254 GCTACGCGCAGCTCAGATGTGAAACAAGAAAGAACTGTGCTGATCTTGTGCTA 1313
DB 1141 GCTACGCGCAGCTCAGATGTGAAACAAGAAAGAACTGTGCTGATCTTGTGCTA 1200
QY 1314 CCGTGAAGTCTCCCTGAGCAATTCACATGAGAGGTTGAGAGGCTTGTGCTGCT 1373
DB 1201 CCGTGAAGTCTCCCTGAGCAATTCACATGAGAGGTTGAGAGGCTTGTGCTGCT 1260
QY 1374 GTGAGCTGAGCAGCTTCTGTAAGTCTGTAAGTCTGTAAGTCTGTAAGTCTGTA 1433
DB 1261 GTGAGCTGAGCAGCTTCTGTAAGTCTGTAAGTCTGTAAGTCTGTAAGTCTGTA 1320
QY 1434 CTACATGCTGTGGGTATCAGCTAGTGTGATGTGTGCTCAACAGTCTCAACAGAG 1493
DB 1321 CTACATGCTGTGGGTATCAGCTAGTGTGATGTGTGCTCAACAGTCTCAACAGAG 1380
QY 1494 GACTTTGTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1553
DB 1381 GACTTTGTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1554 ATCCACAGTGAACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1613
DB 1441 ATCCACAGTGAACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1614 ACTTCAGAGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1673
DB 1501 ACTTCAGAGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1674 TCCATTAACCTCCCTGCTCAGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1733
DB 1561 TCCATTAACCTCCCTGCTCAGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY 1734 CACAGCAATTTTCCACAGAGATGCAAAATGTAAGCCCTTGAAGTCCAGTGAATCA 1793
DB 1621 CACAGCAATTTTCCACAGAGATGCAAAATGTAAGCCCTTGAAGTCCAGTGAATCA 1680
QY 1794 GAAGAGTGTGGAACAGGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1853
DB 1681 GAAGAGTGTGGAACAGGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1854 AATCAGAAATATGAACAAATCTGCAACAAGAGCAAGCTCTTAAGTCAAGGCTG 1913
DB 1741 AATCAGAAATATGAACAAATCTGCAACAAGAGCAAGCTCTTAAGTCAAGGCTG 1800
QY 1914 CCGTGGCTGCTATTAATATCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1973
DB 1801 CCGTGGCTGCTATTAATATCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860

QY 1974 ACCTGTGAAGCTGCCATCTGTTAATCTAAATTCOCATATAG 2017
 DB 1861 ACCTGTGAAGCTGCCATCTGTTAATCTAAATTCOCATATAG 1904

RESULT 3

AA294211
 ID AA294211 standard; cDNA: 2065 BP.

AA294211;

19-JUN-2000 (first entry)

XX Human transferase TRNSFS-11 cDNA clone 2617407CB1.

XX Transferase: TRNSFS-11; human: antitumor; cell proliferation;

XX inflammation; gastrointestinal disorder; developmental disorder;

XX genetic disorder; neurological disorder; reproductive disorder;

XX smooth muscle disorder; immunological disorder; gene therapy;

XX diagnosis; N-acetylglucosamine 6-O-sulfotransferase; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 174..1334

XX FT /*tag= a

XX WC200014251-A2.

XX 16-MAR-2000.

XX 09-SEP-1999: 99NC-US20989.

XX 10-SEP-1998: 98US-0150657.

XX 04-NOV-1998: 98US-0186779.

XX 11-MAY-1999: 99US-0133642.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;

XX Hillman JL, Azimzai Y;

XX WPI: 2000-256996/22.

XX P-PSDB: AAY79219.

XX Human transferase proteins useful for preventing, diagnosing and

XX treating cancers and developmental, gastrointestinal, genetic,

XX immunological, neurological, reproductive and smooth muscle disorders -

XX Claim 9; Page 104-105; 113pp; English.

XX The present sequence is that of cDNA clone 2617407CB1 encoding

XX human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human

XX transferase proteins of the invention (see AAY79209-23). The clone

XX was isolated from gall bladder cDNA library GBLANOT01. TRNSFS-11

XX is expressed in dermatologic and gastrointestinal tissues,

XX especially those associated with inflammation and cell

XX proliferation. It shows homology to mouse N-acetylglucosamine

XX 6-O-sulfotransferase. A polynucleotide comprising nucleotides

XX 264-333 or 1272-1331 of the present sequence can be used as a DNA

XX probe. The new human transferases and polynucleotides can be used

XX in the diagnosis, prevention and treatment (including gene therapy

XX and antisense therapy) of cancer, developmental disorders,

XX gastrointestinal disorders, genetic disorders, immunological

XX disorders, neurological disorders, reproductive disorders, and

XX smooth muscle disorders.

SO Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other;

Query Match

Best Local Similarity 99.8%; Score 1615; DB 21; Length 2065;

Matches 2005; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 9 GCCAGATGCGCTCCAGTCTGGGGAAATGCTTCTCATTTGCTTCCAGCCACCTC 68
 DB 1 GCCAGATGCGCTCCAGTCTGGGGAAATGCTTCTCATTTGCTTCCAGCCACCTC 60
 QY 69 AAGCAGTCTCCCAACCCCTTGAATCTCAGCAGTGTAAAGCTGTACTTTCACAGCTCC 128
 DB 61 AAGCAGTCTCCCAACCCCTTGAATCTCAGCAGTGTAAAGCTGTACTTTCACAGCTCC 120
 QY 129 TGGAGCGAGTGTCTTCTCAGCCCGCTTGCAGAGCTTCCACTTCAGACAGATGCTAC 188
 DB 121 TGGAGCGAGTGTCTTCTCAGCCCGCTTGCAGAGCTTCCACTTCAGACAGATGCTAC 180
 QY 189 TGCCTAAAAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCCATCTTGCTCTAT 248
 DB 181 TGCCTAAAAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCCATCTTGCTCTAT 240
 QY 249 TCTTCCAGATGTACAGCCACACATACATAGCTCCCTGCTATGAAGGACAGCCAGCGCA 308
 DB 241 TCTTCCAGATGTACAGCCACACATACATAGCTCCCTGCTATGAAGGACAGCCAGCGCA 300
 QY 309 TGCACGTGCTGCTTCTGCTTCTTCCGTGCGCTCTGCTCTTCTTGTGGGACACTTTTG 368
 DB 301 TGCACGTGCTGCTTCTGCTTCTTCCGTGCGCTCTGCTCTTCTTGTGGGACACTTTTG 360
 QY 369 GGCAGCACCCAGATGTTTCTTACCTGATGAGCCCGCTGCGACGTGTGATGACCTTCA 428
 DB 361 GGCAGCACCCAGATGTTTCTTACCTGATGAGCCCGCTGCGACGTGTGATGACCTTCA 420
 QY 429 AGCAGACACCCGCTGATGCTGACATGCTGCTGCGGATCTGATACAGGGCGCTCTTC 488
 DB 421 AGCAGACACCCGCTGATGCTGACATGCTGCTGCGGATCTGATACAGGGCGCTCTTC 480
 QY 489 TGTGCGATGAGCGTCTTGTATGCTTACATGAGACCTGTCCCGAGACAGTCCAGCC 548
 DB 481 TGTGCGATGAGCGTCTTGTATGCTTACATGAGACCTGTCCCGAGACAGTCCAGCC 540
 QY 549 TCTTTCAGTGGAGAACAGCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608
 DB 541 TCTTTCAGTGGAGAACAGCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 QY 609 AAGATGAATCATCCCGCGGCTCAGTGCAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 667
 DB 601 AAGATG-AATCATCCCGCGGCTCAGTGCAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 659
 QY 668 GTGTGAGAGAGCGCTGCGCTCTTACAGCAGCAGTGTGTCAAGAGAGTGGCTTCTTC 727
 DB 660 GTGTGAGAGAGCGCTGCGCTCTTACAGCAGCAGTGTGTGTCAAGAGAGTGGCTTCTTC 719
 QY 728 AACCTGAGTCCCTTACCCGCTGCTGAGAAAGACCCCTCCCTCAACTGATATCGGCAC 787
 DB 720 AACCTGAGTCCCTTACCCGCTGCTGAGAAAGACCCCTCCCTCAACTGATATCGGCAC 779
 QY 788 CTGTGCGGAGACCCCGGCGCTGCTTCCGCTTCCGAGAAAGCAAGAGAGATCTCATG 847
 DB 780 CTGTGCGGAGACCCCGGCGCTGCTTCCGCTTCCGAGAAAGCAAGAGAGATCTCATG 839
 QY 848 ATTGACAGTGTGATGATGGGAGAGATGAGCAAAACTCAAGAGAGGAGCAAAACC 907
 DB 840 ATTGACAGTGTGATGATGGGAGAGATGAGCAAAACTCAAGAGAGGAGCAAAACC 899
 QY 908 TACTATGTATGAGTATGCTCCCAAGCCAGCTGAGATCTTCAAGACATCATCAGTCC 967
 DB 900 TACTATGTATGAGTATGCTCCCAAGCCAGCTGAGATCTTCAAGACATCATCAGTCC 959
 QY 968 TTGCCCCAAGGCGCTGAGAGAGAGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1027
 DB 960 TTGCCCCAAGGCGCTGAGAGAGAGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019
 QY 1028 CCTGTGCGCCAGACTTCCGAATGTATGATGATGATGATGATGATGATGATGATGATGAT 1087
 DB 1020 CCTGTGCGCCAGACTTCCGAATGTATGATGATGATGATGATGATGATGATGATGATGAT 1079
 QY 1088 CAACTGAGTGTATTAACATCAACCCAGGCAAGGAGATGAGTACCAAGCTTTTCCACACA 1147

```

Db 1080 CAGACCTGGGTGCATACATCACCCAGAGGAGGAGGATGGGTACACAGCTTTCCACACA 1139
Qy 1148 AATGCCAGGAGATGCCCTTAATGTCTCCAGGCTGGCGCTGTCTTCCCTATGAAGAAG 1207
Db 1140 AATGCCAGGAGATGCCCTTAATGTCTCCAGGCTGGCGCTGTCTTCCCTATGAAGAAG 1199
Qy 1208 GTTCTCGAGCTTCAGAAAGGCTGTGGGATGCCATGAAATTTGCTGGCTACCCGCCAGTC 1267
Db 1200 GTTCTCGAGCTTCAGAAAGGCTGTGGGATGCCATGAAATTTGCTGGCTACCCGCCAGTC 1259
Qy 1268 AGATCTGACAGAGAGAGAAACCTGTGCTGATCTTCTGCTACCTGAGCTGTCCCT 1327
Db 1260 AGATCTGACAGAGAGAGAAACCTGTGCTGATCTTCTGCTACCTGAGCTGTCCCT 1319
Qy 1328 GAGCAAAATCCACTAGAGGGTTGAGAAAGGCTTGTGCTGCCACCTGTGTCTACGCTCA 1387
Db 1320 GAGCAAAATCCACTAGAGGGTTGAGAAAGGCTTGTGCTGCCACCTGTGTCTACGCTCA 1379
Qy 1388 CTTCCTCTGATGATGCTTGTGAGCCCTGGCTACATCTCTGAGCCTTAACATACATCTGTG 1447
Db 1380 CTTCCTCTGATGATGCTTGTGAGCCCTGGCTACATCTCTGAGCCTTAACATACATCTGTG 1439
Qy 1448 GTATCACACTGAGTGTGAGTTGTGTCCACAGTGTCTCAGCAGAGAGACTTTTGTGTCCA 1507
Db 1440 GTATCACACTGAGTGTGAGTTGTGTCCACAGTGTCTCAGCAGAGAGACTTTTGTGTCCA 1499
Qy 1508 TGTCTGTGTCTAGAAACAGAGCTGGGAGACCTTATGTGACAGACAGATCCACAGAGGAA 1567
Db 1500 TGTCTGTGTCTAGAAACAGAGCTGGGAGACCTTATGTGACAGACAGATCCACAGAGTGA 1559
Qy 1568 ACAGGGATATGCTCTCTCTTCTTCTTCTGATCTCTCTGCTGTCTGGGAGAGCTTCAGAGACT 1627
Db 1560 ACAGGGATATGCTCTCTCTTCTTCTTCTGATCTCTCTGCTGTCTGGGAGAGCTTCAGAGACT 1619
Qy 1628 GTGGCCTGGAGGCTTATTAGACAGACAGATGATGATGATGATGATGATGATGATGATGAT 1687
Db 1620 GTGGCCTGGAGGCTTATTAGACAGACAGATGATGATGATGATGATGATGATGATGATGAT 1679
Qy 1688 TGTCCACATCTGTCACCAATGGGAGATGATGATGATGATGATGATGATGATGATGATGAT 1747
Db 1680 TGTCCACATCTGTCACCAATGGGAGATGATGATGATGATGATGATGATGATGATGATGAT 1739
Qy 1748 ACAGAGATGCAAAATTCGAGGCTTGGAGTTCCAGTGTGATTCAGAGAGAGAGAGTGGGAA 1807
Db 1740 ACAGAGATGCAAAATTCGAGGCTTGGAGTTCCAGTGTGATTCAGAGAGAGAGAGTGGGAA 1799
Qy 1808 CAAAGTGGATGCTCTTATGAGCTTGAAGCTGACATGACAGCTATGCTATGAGAAATATGA 1867
Db 1800 CAAAGTGGATGCTCTTATGAGCTTGAAGCTGACATGACAGCTATGCTATGAGAAATATGA 1859
Qy 1868 AACAAATCTCTGACAAAAGAGCAAGCTCTTAAGTTACAGAGGCTGCTGGCTGATTT 1927
Db 1860 AACAAATCTCTGACAAAAGAGCAAGCTCTTAAGTTACAGAGGCTGCTGGCTGATTT 1919
Qy 1928 GAATATGATCTGCTGCTGATTTTCCATGACATAGAAAGCTTGAAGCTGAGTGC 1987
Db 1920 GAATATGATCTGCTGCTGATTTTCCATGACATAGAAAGCTTGAAGCTGAGTGC 1979
Qy 1988 CATCTGTTAACTAAATTCACCAATATAG 2017
Db 1980 CATCTGTTAACTAAATTCACCAATATAG 2009

```

```

RESULT 4
AAS16947
ID AAS16947 standard; cDNA; 1333 BP.
AC XX
XX AAS16947;
XX
XX 12-MAR-2002 (first entry)
XX
XX Human L-selectin sulfoltransferase-2 (LST-2) cDNA.
DE

```

```

XX Human; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
KW L-selectin; L-selectin sulfoltransferase-2; Crohn's disease; diabetes;
KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;
KW antileukemia; antiproliferative; antidiabetic; dermatological;
KW antiallergic.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 111..1253
XX FT /*tag= a
XX FT /product= "Human LST-2"
XX
XX WO200185177-A1.
XX
XX 15-NOV-2001.
XX
XX 10-MAY-2001; 2001WO-US15452.
XX
XX 11-MAY-2000; 2000US-0569320.
XX
XX (BURN-) BURNHAM INST.
XX
XX Fukuda M, Yeh J, Hirooka N;
XX
XX MPI: 2002-075226/10.
XX P-PSDB: MAU11274.
XX
XX New enzyme, useful for modifying acceptor molecule, comprises an
XX isolated L-selectin sulfoltransferase-2 that directs expression of
XX L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
XX intestinal GLC6C 6-sulfoltransferase
XX
XX Claim 19; Fig 4; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor molecule
XX by contacting the acceptor with an isolated
XX beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active
XX fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The
XX invention also provides a method of treating or preventing an
XX L-selectin-mediated condition by reducing the expression or activity of a
XX beta1,3GNT that directs expression of a MECA-79 antigen. This can be done
XX by administering to the subject an oligosaccharide L-selectin antagonist
XX that inhibits binding of L-selectin to a MECA-79 antigen, for example by
XX administering antibody material that specifically binds beta1,3GNT,
XX and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin
XX sulfoltransferase-2 (LST-2) also directs MECA-79 antigen expression.
XX Alternatively, the expression or activity of LST-2 or its active
XX fragment can be reduced in combination with reducing the expression or
XX activity of beta1,3GNT. The method is useful for treating L-selectin
XX mediated conditions such as Crohn's disease and ulcerative colitis,
XX inflammatory disorders of the skin such as allergic contact dermatitis,
XX psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
XX hypersensitivity reactions, diabetes and hyperplastic thymus. This
XX sequence represents cDNA encoding human LST-2.
XX
XX Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other:
XX
XX Query Match 59.6%; Score 1211; DB 24; Length 1333;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1261; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 161 AAGGTCTCCACTGACGACCAATGCTACTGCTAATAAAATGAGCTCTGCTGTTCTG 220
Db 72 AAGGTCTCCACTGACGACCAATGCTACTGCTAATAAAATGAGCTCTGCTGTTCTG 131
Qy 221 GTTCCCAATGAGGCAATGCTGCTATTTCTGACAGATGACGACCAATGAGCTCC 280
Db 132 GTTCCCAATGAGGCAATGCTGCTATTTCTGACAGATGACGACCAATGAGCTCC 191

```

```
QY 281 CTGTCTATGAGGACACGCGGAGGATGACGTGCTGTTCTGTCTCTCTGCGGCTCT 340
    |||||||
Db 192 CTGTCTATGAGGACACGCGGAGGATGACGTGCTGTTCTGTCTCTCTGCGGCTCT 251
QY 341 GGCCTCTCTTTTGTGGGACGCTTTTGGGACGACCCAGATGTTTCTACTGATGAG 400
    |||||||
Db 252 GGCCTCTCTTTTGTGGGACGCTTTTGGGACGACCCAGATGTTTCTACTGATGAG 311
QY 401 CCCGCTGAGCAGTGTGATACCTTCAAGCAGACACCGCTGATGCTGACATGGCT 460
    |||||||
Db 312 CCCGCTGAGCAGTGTGATACCTTCAAGCAGACACCGCTGATGCTGACATGGCT 371
QY 461 GTGGGGATCTGATACGCGGCGCTTCTTGTGACATGACGCTTGTGATGCTTACATG 520
    |||||||
Db 372 GTGGGGATCTGATACGCGGCGCTTCTTGTGACATGACGCTTGTGATGCTTACATG 431
QY 521 GAACCTGTGCTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAGAAACAGCCGGGCGCTGTGT 580
    |||||||
Db 432 GAACCTGTGCTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAGAAACAGCCGGGCGCTGTGT 491
QY 581 TCTGACCTGCTGCTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGACAG 640
    |||||||
Db 492 TCTGACCTGCTGCTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGACAG 551
QY 641 CTCCTGTGCACTCAACAGCCCTTGTAGGTGTGAGAGAGGCTGCTGCTTACAGCCAC 700
    |||||||
Db 552 CTCCTGTGCACTCAACAGCCCTTGTAGGTGTGAGAGAGGCTGCTGCTTACAGCCAC 611
QY 701 GTGGTGTCAAGAGAGGTCGCTTCTTCACTGATGCTGCTCTTACCCGCTGTGAAAGAC 760
    |||||||
Db 612 GTGGTGTCAAGAGAGGTCGCTTCTTCACTGATGCTGCTCTTACCCGCTGTGAAAGAC 671
QY 761 CCCCTCCCACTGACATATGCTGACCTGTGCTCGGAGACCCCGGCGGCTGCTGCTTCC 820
    |||||||
Db 672 CCCCTCCCACTGACATATGCTGACCTGTGCTCGGAGACCCCGGCGGCTGCTGCTTCC 731
QY 821 CGAAGAGGACAAAGGAGATCTCTATGATGACAGTGCATTTGTATGGGAGCATGAG 880
    |||||||
Db 732 CGAAGAGGACAAAGGAGATCTCTATGATGACAGTGCATTTGTATGGGAGCATGAG 791
QY 881 CAGAAGCTCAAGAGAGGACCAACCTACTATGATGATGACAGTGCATCTGCCAAGCCAG 940
    |||||||
Db 792 CAGAAGCTCAAGAGAGGAGACCAACCTACTATGATGATGACAGTGCATCTGCCAAGCCAG 851
QY 941 CTGAGAGTCTCAAGACCATCATGCTTGTGCCCAAGGCCCTGACGAGAGCGTACTGCTT 1000
    |||||||
Db 852 CTGAGAGTCTCAAGACCATCATGCTTGTGCCCAAGGCCCTGACGAGAGCGTACTGCTT 911
QY 1001 GTGGGCTATGAGGAGCTGCTGCGAGCCCTGTGGCCAGACTTCCCAATGTATGAAATTC 1060
    |||||||
Db 912 GTGGGCTATGAGGAGCTGCTGCGAGCCCTGTGGCCAGACTTCCCAATGTATGAAATTC 971
QY 1061 GTGGGATGGAATCTTGTCCCATCTTTCAGACCTGGGTGATACATCACCCGAGGCAAG 1120
    |||||||
Db 972 GTGGGATGGAATCTTGTCCCATCTTTCAGACCTGGGTGATACATCACCCGAGGCAAG 1031
QY 1121 GGCATGGGTGACCAAGCTTTTCCACAAATGCCAGGATGCCCTTATATGTCTCCCAAGCT 1180
    |||||||
Db 1032 GGCATGGGTGACCAAGCTTTTCCACAAATGCCAGGATGCCCTTATATGTCTCCCAAGCT 1091
QY 1181 TGGGCGTGTCTTTGCCCTATGAAAAGTTTTCGACTTGAAGAAGCCTTGGGATGCC 1240
    |||||||
Db 1092 TGGGCGTGTCTTTGCCCTATGAAAAGTTTTCGACTTGAAGAAGCCTTGGGATGCC 1151
QY 1241 ATGATTTTGTGGGCTACCGCCAGTCAAGATCTGAACAAGACAGAAACCTGTTGCTG 1300
    |||||||
Db 1152 ATGATTTTGTGGGCTACCGCCAGTCAAGATCTGAACAAGACAGAAACCTGTTGCTG 1211
QY 1301 GATTTTGTCTTACTGAGTGTGCTGACCAATCCACTAAGAGGTTTGAAGAGCTTT 1360
    |||||||
Db 1212 GATTTTGTCTTACTGAGTGTGCTGACCAATCCACTAAGAGGTTTGAAGAGCTTT 1271
QY 1361 GCTGCCACCTGAGTGTCAAGCCTCACTTCTCTGAATGCTCTGAGCCTTGCTTACAT 1420
```

```
Db 1272 GGTGCACACCTGTGGTGAAGCCTCAGTCACTTCTCTGAATGCTTGTAGGCTTGCTTACAT 1331
    |||||||
QY 1421 CT 1422
    ||
Db 1332 CT 1333

RESULT 5
AAK91803
ID AAK91803 standard; cDNA: 877 BP.
XX
AC AAK91803;
XX
DT 06-NOV-2001 (first entry)
DE Human cDNA 5'-end sequence, SEQ ID NO: 263.
DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI- ) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 36.5%; Score 741; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 1,1e-305;
Matches 741; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 ACTTTCACAGCTTCTCGGAGCGAGTCTTTCAGACCCGCTTTCGAAGTCTTCCACT 173
    |||||||
Db 1 ACTTTCACAGCTTCTCGGAGCGAGTCTTTCAGACCCGCTTTCGAAGTCTTCCACT 60
QY 174 TCAGCAATGCTACTGCTTAATAAATAAGAGTCTGCTGTTTGGTTCGAGTGG 233
    |||||||
Db 61 TCAGCAATGCTACTGCTTAATAAATAAGAGTCTGCTGTTTGGTTCGAGTGG 120
QY 234 CCATCTGGCTATTTCTCCACATGTACAGCCACAAATCAGCTTCCGTCTATGAGG 293
    |||||||
```

DB 121 CCATCTTGCTCTATCTTCCACATGATACACCAATCATGCTCCCTGCTATGAAG 180
QY 294 CACAGCCCGGAGGATGACGATGCTGCTCTCTCTGCGGCTCTGCTCTCTTTG 353
DB 181 CACAGCCCGGAGGATGACGATGCTGCTCTCTCTGCGGCTCTGCTCTCTTTG 240
QY 354 TGGGGAGCTTTTGGGAGCACCAGATGTTTCTACCTGATGAGACCCGCTGGCAG 413
DB 241 TGGGGAGCTTTTGGGAGCACCAGATGTTTCTACCTGATGAGACCCGCTGGCAG 300
QY 414 TGTGATGATACCTTCAGGAGAGACCGGCTGATGCTGACATGGGTGGGATCTGA 473
DB 301 TGTGATGATACCTTCAGGAGAGACCGGCTGATGCTGACATGGGTGGGATCTGA 360
QY 474 TACGGGCGCTCTCTTGTGCGACATGAGCGTCTTGTGATGCTGACATGAGACTGTCC 533
DB 361 TACGGGCGCTCTCTTGTGCGACATGAGCGTCTTGTGATGCTGACATGAGACTGTCC 420
QY 534 GAGAGATGTCAGCTCTCTTGTGCGACATGAGCGTCTTGTGATGCTGACATGAGACTGTCC 593
DB 421 GAGAGATGTCAGCTCTCTTGTGCGACATGAGCGTCTTGTGATGCTGACATGAGACTGTCC 480
QY 594 GTGACATCATCCCAAGATGAAATCATCCCGGGCTGCTGACAGCTCTGTCAGATC 653
DB 481 GTGACATCATCCCAAGATGAAATCATCCCGGGCTGCTGACAGCTCTGTCAGATC 540
QY 654 AACAGCCCTTTGAGTGTGTGAGAGAGAGCTGCGCTCTCTACAGCCAGCTGTCAAG 713
DB 541 AACAGCCCTTTGAGTGTGTGAGAGAGAGCTGCGCTCTCTACAGCCAGCTGTCAAG 600
QY 714 AGGAGCGCTCTTCAACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 773
DB 601 AGGAGCGCTCTTCAACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 774 TGCATATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 833
DB 661 TGCATATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 834 AGGAGATGCTGATGATGACA 854
DB 721 AGGAGATGCTGATGATGACA 741

RESULT 6
AAK93921
ID AAK93921 standard; cDNA; 877 BP.
AC AAK93921;
XX 06-NOV-2001 (first entry)
DE Human cDNA clone representative sequence, SEQ ID NO: 2381.
XX Human: full length cDNA; cDNA synthesis; oligo-capping; ss.
KM Homo sapiens.
OS Homo sapiens.
PN EP1130094-A2.
PD 05-SEP-2001.
XX 07-JUL-2000; 2000EP-0114089.
XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX (HELI-) HELIX RES INST.
PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS
XX Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' and 3' ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used in
CC representative sequence to identify the clone.
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
Query Match 36.5%; Score 741; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 1.1e-305;
Matches 741; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 114 ACTTTCACAGCTTCCTGAGAGAGAGTGTCTCTCAAGCCGCTTGAAGCTCTCCACT 173
DB 1 ACTTTCACAGCTTCCTGAGAGAGAGTGTCTCTCAAGCCGCTTGAAGCTCTCCACT 60
QY 174 TCAGCAATGCTACTGCT 233
DB 61 TCAGCAATGCTACTGCT 120
QY 234 CCATCTTGCTCTATCTTCCACATGATGACGACCAATCATGAGCTGCTGCTGCTGCTGCTGCT 293
DB 121 CCATCTTGCTCTATCTTCCACATGATGACGACCAATCATGAGCTGCTGCTGCTGCTGCTGCT 180
QY 294 CACAGCCCGGAGGATGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 353
DB 181 CACAGCCCGGAGGATGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 354 TGGGGAGCTTTTGGGAGCACCAGATGTTTCTACCTGATGAGACCCGCTGGCAG 413
DB 241 TGGGGAGCTTTTGGGAGCACCAGATGTTTCTACCTGATGAGACCCGCTGGCAG 300
QY 414 TGTGATGATACCTTCAGGAGAGACCGGCTGATGCTGACATGGGTGGGATCTGA 473
DB 301 TGTGATGATACCTTCAGGAGAGACCGGCTGATGCTGACATGGGTGGGATCTGA 360
QY 474 TACGGGCGCTCTCTTGTGCGACATGAGCGTCTTGTGATGCTGACATGAGACTGTCC 533
DB 361 TACGGGCGCTCTCTTGTGCGACATGAGCGTCTTGTGATGCTGACATGAGACTGTCC 420
QY 534 GAGAGATGTCAGCTCTCTTGTGCGACATGAGCGTCTTGTGATGCTGACATGAGACTGTCC 593
DB 421 GAGAGATGTCAGCTCTCTTGTGCGACATGAGCGTCTTGTGATGCTGACATGAGACTGTCC 480
QY 594 GTGACATCATCCCAAGATGAAATCATCCCGGGCTGCTGACAGCTCTGTCAGATC 653
DB 481 GTGACATCATCCCAAGATGAAATCATCCCGGGCTGCTGACAGCTCTGTCAGATC 540
QY 654 AACAGCCCTTTGAGTGTGTGAGAGAGAGCTGCGCTCTCTACAGCCAGCTGTCAAG 713
DB 541 AACAGCCCTTTGAGTGTGTGAGAGAGAGCTGCGCTCTCTACAGCCAGCTGTCAAG 600
QY 714 AGTGGCGCTCTTCAACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 773
DB 601 AGTGGCGCTCTTCAACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 774 TGCATATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 833
DB 661 TGCATATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720

QY 834 AGGAGATCTCATGATTGACA 854
|||||
Db 721 AGGAGATCTCATGATTGACA 741

RESULT 7

ABK547724
ID ABK54724 standard; cDNA: 517 BP.

AC ABK54724:

DT 18-JUN-2002 (first entry)

XX Human colon cancer-associated cDNA, SEQ ID NO 194.

XX Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.

XX Homo sapiens.

XX WO200212280-A2.

XX 14-FEB-2002.

XX 30-JUL-2001; 2001WO-US23826.

XX 03-AUG-2000; 2000US-223265P.

XX 02-OCT-2000; 2000US-237406P.

XX 20-MAR-2001; 2001US-277495P.

XX 03-JUL-2001; 2001US-302702P.

XX (CORI-) CORIXA CORP.

XX Pyle RA, Xu J, Secrist H;

XX WPI; 2002-257462/30.

XX Novel polynucleotide encoding colon tumour polypeptides, useful as

XX vaccines for treating colon cancers -

XX Claim 1; Page 206; 425pp; English.

XX The invention relates to isolated polynucleotides (i) encoding colon

XX tumour polypeptides (ii). (i) is useful for stimulating an immune

XX response in a patient and treating colon cancer in a patient.

XX Oligonucleotides derived from (i) are useful for determining the presence

XX of cancer in a patient. (i) and (ii) are useful in pharmaceutical

XX compositions, e.g. vaccines, and other compositions for the diagnosis

XX and treatment of colon cancer. A composition comprising a first component

XX selected from physiologically acceptable carriers and immunostimulants,

XX and an antigen-presenting cell expressing (ii) is useful for inhibiting

XX development of cancer in a patient. (i) is useful in the design and

XX preparation of ribozyme molecules for inhibiting expression of tumour

XX polypeptides and (i). ABK54531-ABK5464 represent human colon cancer cDNA

XX sequences of the invention.

XX Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

XX Query Match 24.9%; Score 505; DB 24; Length 517;

XX Best Local Similarity 100.0%; Pred. No. 3.8e-205;

XX Matches 505; Conservative 0; Mismatches 0; Idels 0; Gaps 0;

QY 1355 GCGTTTGCTGCCACCTGGTGCACGCTCAGTCACTTCTCGAATGCTTCTGAGCCTTGC 1414
|||||
Db 13 GCGTTTGCTGCCACCTGGTGCACGCTCAGTCACTTCTCGAATGCTTCTGAGCCTTGC 72
QY 1415 CTACATCTCTAGGCTTAACATACATGCTGTGGGTATCACACTGAGTGTGAGTGTGCC 1474
|||||
Db 73 CTACATCTCTAGGCTTAACATACATGCTGTGGGTATCACACTGAGTGTGAGTGTGCC 132
QY 1475 ACACGTGCTCAAGAGAAGAACTTTGTGTCCATGCTTGTGTCTTAGAAAACAGACTGGGG 1534
|||||
Db 133 ACACGTGCTCAAGAGAAGAACTTTGTGTCCATGCTTGTGTCTTAGAAAACAGACTGGGG 192

QY 1535 AACCTTATGTGAGCAGACACATCCACCACTGAACACAGGATATTGCTCTCTTTTCTT 1594
|||||
Db 193 AACCTTATGTGAGCAGACACATCCACCACTGAACACAGGATATTGCTCTCTTTTCTT 252
QY 1595 GATCTTCCTGCTGGGAGAGCTTCAGAGACTTTGGGCGTGGAGGCTTATTAAGCAGCAG 1654
|||||
Db 253 GATCTTCCTGCTGGGAGAGCTTCAGAGACTTTGGGCGTGGAGGCTTATTAAGCAGCAG 312
QY 1655 ACAGTATCACTGGAATGATTCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGGAATG 1714
|||||
Db 313 ACAGTATCACTGGAATGATTCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGGAATG 372
QY 1715 GATCTTCACCAAGAGCTTCACGAGCATTTTCCACAGAGATGCAATTTCTGAGCCTTGG 1774
|||||
Db 373 GATCTTCACCAAGAGCTTCACGAGCATTTTCCACAGAGATGCAATTTCTGAGCCTTGG 432
QY 1775 AGTTCACAGTGAATTCAGAGGAAGTGGGAACAAGTTGATGGCTACTATGAGCTT 1834
|||||
Db 433 AGTTCACAGTGAATTCAGAGGAAGTGGGAACAAGTTGATGGCTACTATGAGCTT 492
QY 1835 GACCATCACAGCTATCGGTATCAG 1859
|||||
Db 493 GACCATCACAGCTATCGGTATCAG 517

RESULT 8

AAK92588/C
ID AAK92588 standard; cDNA: 548 BP.

AC AAK92588:

XX 06-NOV-2001 (first entry)

XX Human cDNA 3'-end sequence, SEQ ID NO: 1048.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

XX 830 Primers useful for synthesizing full length cDNA clones and their

XX use in genetic manipulation -

XX Claim 3; SEQ ID NO 1048; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA

XX clones. 830 cDNA molecules encoding a human protein have been

XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

XX molecules have been determined. Primers for synthesizing the full length

XX cDNA are useful for clarifying the function of the protein encoded by

XX the cDNA. The full length clones were obtained by construction of full

XX length enriched cDNA libraries that were synthesized by the oligo-capping

XX method. The primers enable the production of the full length cDNA easily

XX without any special methods. The present sequence is the nucleotide

XX sequence of the 3'-end of a cDNA provided in the invention.

XX Note: The sequence data for this patent did not form part of the printed

CC Specification, but was obtained in CD-ROM format directly from EPO.
XX Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;
SQ
Query Match 17.3%; Score 351; DB 22; Length 548;
Best Local Similarity 100.0%; Pred. No. 1.5e-139; Mismatches 0; Gaps 0;
Matches 351; Conservative 0; Indels 0; Gaps 0;
QY 1667 GAATTGATCCATAAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCTTTCACCA 1726
DB 426 GAATTGATCCATAAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCTTTCACCA 367
QY 1727 AAGAGCTCACCAGATTTTCCACAGATGCAATCTTGAGCCCTTGAGATTTCCAGTGG 1786
DB 366 AAGAGCTCACCAGATTTTCCACAGATGCAATCTTGAGCCCTTGAGATTTCCAGTGG 307
QY 1787 ATTCAAGGAAGAGAGTGGAGACAAGTTGGATGCTTACTATAGCTTACCATCAGC 1846
DB 306 ATTCAAGGAAGAGAGTGGAGACAAGTTGGATGCTTACTATAGCTTACCATCAGC 247
QY 1847 TATCGGTATACAGAAATATGAAACAAATCTTCGACAAAAGAGCAAGCTTTAAGTTCA 1906
DB 246 TATCGGTATACAGAAATATGAAACAAATCTTCGACAAAAGAGCAAGCTTTAAGTTCA 187
QY 1907 CAGGCTCCTGGGCTGCATTTGATATACCTTCCTCTGCATTTTCCATCATGAA 1966
DB 186 CAGGCTCCTGGGCTGCATTTGATATACCTTCCTCTGCATTTTCCATCATGAA 127
QY 1967 GACTTTGACCTGTGAGACCTGCTGTGTATACCTTAATCCCAATTAAG 2017
DB 126 GACTTTGACCTGTGAGACCTGCTGTGTATACCTTAATCCCAATTAAG 76
RESULT 9
AAC76156
ID AAC76156 standard; cDNA; 2988 BP.
AC AAC76156;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.
XX
Human: open reading frame; ORFX; detection: cytostatic; hepatotropic;
KM vulnery; antiporiatic; antiparkinsonian; nocrotropic; neuroprotective;
KM anticonvulsant; osteopathic; antirachitic; immunosuppressant; cardiac;
KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
KM antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
KM antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KM cholesterol ester storage; systemic lupus erythematosus; infection;
KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
KM thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN MO200058473-A2.
XX
PD 05-OCT-2000.
XX
PE 31-MAR-2000; 2000MO-US08621.
XX
PR 31-MAR-1999; 9905-0127607.
PR 02-APR-1999; 9905-0127636.
PR 05-APR-1999; 9905-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach M;
XX WPI: 2000-602362/57.
DR P-PSDB; AAB41947.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
Claim 5; Page 2597-2599; 5507pp; English.
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiporiatic; antiparkinsonian; nocrotropic; neuroprotective;
CC osteopathic; anticonvulsant; antirachitic; immunosuppressant;
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antihypoid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;
Query Match 15.1%; Score 306; DB 21; Length 2988;
Best Local Similarity 100.0%; Pred. No. 2e-120;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 506 TTGATGCTACATGAACCTGCTCCCGGAGACAGCTCCAGCTTTCAGTGGAGAGAC 565
DB 1480 TTGATGCTACATGAACCTGCTCCCGGAGAGCTCCAGCTTTCAGTGGAGAGAC 1539
QY 566 AGCCGGGCCCTGTGTTGACACCTGCTGTGACATCATCCACAAGATGAATCATCCCC 625
DB 1540 AGCCGGGCCCTGTGTTGACACCTGCTGTGACATCATCCACAAGATGAATCATCCCC 1599
QY 626 CGGCTCACTGAGGCTCCTGTGACGACAGCCCTTGAAGGAGTGAAGAAGGCTGC 685
DB 1600 CGGCTCACTGAGGCTCCTGTGACGACAGCCCTTGAAGGAGTGAAGAAGGCTGC 1659
QY 686 CGCTCTACAGCAGCTGTGCTCAAGAGAGTGCCTTCTCAACCTGACAGTCCCTTAC 745
DB 1660 CGCTCTACAGCAGCTGTGCTCAAGAGAGTGCCTTCTCAACCTGACAGTCCCTTAC 1719
QY 746 CCGCTCTGAAGAGCCCTCCTCAACTGATATCGAGCACTGTGCGGAGACCCCGG 805
DB 1720 CCGCTCTGAAGAGCCCTCCTCAACTGATATGAGCACTGTGCGGAGACCCCGG 1779
QY 806 GCCGTG 811
DB 1780 GCCGTG 1785
RESULT 10
ABK54794
ID ABK54794 standard; cDNA; 389 BP.
AC ABK54794;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human colon cancer-associated cDNA, SEQ ID NO 264.

XX
PI Rosen SD, Lee JK, Hemmerich S;
XX
DR WPI: 2001-138471/14.
DR P-PSDB: AAT72640.
XX
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications
XX
PS Claim 6; Fig 4A; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
CC beta) cDNA. GST-4 gene is found on chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation.
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
XX
SQ Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;
Query Match 2.9%; Score 59; DB 22; Length 1694;
Best Local Similarity 100.0%; Pred. No. 3.3e-15;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 678 AGGCGTGGCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTTCACCTGCAG 736
DB 676 AGGCGTGGCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTTCACCTGCAG 734
RESULT 14
AADD02699 standard; cDNA: 2044 BP.
ID AADD02699 standard; cDNA: 2044 BP.
XX
AC AADD02699;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.
XX
KW Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1; ss.
XX
CS Homo sapiens.
XX
FH Key Location/Qualifiers
FH CDS 218..1390
FT /*tag= a
FT /product= "Human glycosyl sulfotransferase-4alpha
FT (GST-4alpha)"
FT /note= "CDS is specifically claimed as SEQ ID NO: 4
FT in claim 6 (page no: 41) of the specification"
XX
XX W0200106015-A1.

PD 25-JAN-2001.
XX
XX 19-JUL-2000; 200GWO-US19741.
PF
XX 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX
XX (REGC) UNIV CALIFORNIA.
XX
PI Rosen SD, Lee JK, Hemmerich S;
XX
DR WPI: 2001-138471/14.
DR P-PSDB: AAT72639.
XX
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications
XX
PS Claim 6; Fig 1; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
CC alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation.
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.
XX
SQ Sequence 2044 BP; 422 A; 645 C; 570 G; 407 T; 0 other;
Query Match 2.9%; Score 59; DB 22; Length 2044;
Best Local Similarity 100.0%; Pred. No. 3.3e-15;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 678 AGGCGTGGCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTTCACCTGCAG 736
DB 708 AGGCGTGGCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCGCTTTCACCTGCAG 766
RESULT 15
AADD02698 standard; cDNA: 2170 BP.
ID AADD02698 standard; cDNA: 2170 BP.
XX
AC AADD02698;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.
XX
KW Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1; ss.
XX
CS Homo sapiens.
XX
FH Key Location/Qualifiers
FH CDS 218..1390
FT /*tag= a
FT /product= "Human glycosyl sulfotransferase-4alpha
FT (GST-4alpha)"
FT /note= "CDS is specifically claimed as SEQ ID NO: 4
FT in claim 6 (page no: 41) of the specification"
XX
XX W0200106015-A1.

```

FH Key Location/Qualifiers
FT 5'UTR 9..343
FT CDS /*tag= a
FT 344..1516
FT /*tag= b
FT /*product= "Human glycosyl sulfotransferase-4alpha
FT (GST-4alpha)"
FT /*note= CDS is specifically claimed as SEQ ID NO: 4
FT in claim 6 (page no: 41) of the specification"
FT 1517..2134
FT /*tag= c
FT 3'UTR
XX
XX WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-DS19741.
XX
XX 20-JUL-1999; 99US-0144694.
XX 13-JUL-2000; 2000US-0593828.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerlich S;
XX
XX WPI: 2001-138471/14.
XX P-PSDB: AAY72639.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4delta and GST-6 for
XX diagnostic and therapeutic agent screening applications
XX
XX Claim 6; Page 62; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX selectin with a non-sulphated selectin ligand, GST and a small molecular
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation.
XX CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adenallitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX
XX
XX Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other;
XX
XX Query Match 2.9%; Score 59; DB 22; Length 2170;
XX Best Local Similarity 100.0%; Pred. No. 3.3e-15;
XX Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 678 AGGCTGCGCGCTCCACAGCCACGCTGCTCTCAAGAGTGCGCTTCTCAACCTGCAG 736
XX |||||||
XX Db 834 AGGCTGCGCGCTCCACAGCCACGCTGCTCTCAAGAGTGCGCTTCTCAACCTGCAG 892
XX
XX RESULT 16
XX ABN89506
XX ID ABN89506 standard; cDNA; 2544 BP.
XX
XX ABN89506;
XX
XX 05-SEP-2002 (first entry)
XX
XX Human corneal N-acetylglucosamine-6-sulfotransferase cDNA SEQ ID NO:1.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; G1CNA6ST;
XX

```

```

KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological; chromosome 16q22; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 693..1880
XX /*tag= a
XX /*product= "N-acetylglucosamine-6-sulfotransferase"
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-0927602.
XX
XX 11-AUG-2000; 2000US-325773P.
XX
XX (FUKU/) FUKUDA M. N.
XX (AKAM/) AKAMA T. O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI: 2002-507643/54.
XX P-PSDB: ABB81554.
XX
XX New nucleic acid encoding corneal
XX N-acetylglucosamine-6-sulfotransferase, useful for treatment,
XX monitoring and diagnosis of macular corneal dystrophy
XX
XX Claim 4; Fig 1A-D; 69pp; English.
XX
XX The present sequence encodes human corneal
XX N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyse
XX sulfation of keratan sulfate (KS). Also described is a method for
XX monitoring the effect of treatments for macular corneal dystrophy (MCD),
XX and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
XX and has ophthalmological activity. (I) can be used to treat or prevent
XX macular corneal dystrophy types I or II. (I) makes possible treatment
XX of MCD without requiring keratoplasty or keratectomy.
XX
XX
XX Sequence 2544 BP; 460 A; 800 C; 733 G; 551 T; 0 other;
XX
XX Query Match 2.9%; Score 59; DB 24; Length 2544;
XX Best Local Similarity 100.0%; Pred. No. 3.3e-15;
XX Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 678 AGGCTGCGCGCTCCACAGCCACGCTGCTCTCAAGAGTGCGCTTCTCAACCTGCAG 736
XX |||||||
XX Db 1180 AGGCTGCGCGCTCCACAGCCACGCTGCTCTCAAGAGTGCGCTTCTCAACCTGCAG 1238
XX
XX RESULT 17
XX ABN89533
XX ID ABN89533 standard; DNA; 48436 BP.
XX
XX ABN89533;
XX
XX 05-SEP-2002 (first entry)
XX
XX Human corneal N-acetylglucosamine-6-sulfotransferase DNA SEQ ID NO:36.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; G1CNA6ST;
XX corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX ophthalmological; chromosome 16q22; gene; ds.
XX
XX Homo sapiens.
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-0927602.
XX

```

| | | | |
|----|----|---|--------------|
| XX | FT | exon | 45094..45185 |
| XX | FT | /tag= e | |
| XX | FT | /number= 3 | |
| XX | FT | /label= 4a-5U2 | |
| XX | FT | 45186..46633 | |
| XX | FT | /tag= f | |
| XX | FT | /cons_splice= (5'site:NO, 3'site:NO) | |
| XX | FT | 46634..46700 | |
| XX | FT | /tag= g | |
| XX | FT | /number= 4 | |
| XX | FT | /label= 4a-5U1 | |
| XX | FT | 46701..47938 | |
| XX | FT | /tag= h | |
| XX | FT | /cons_splice= (5'site:YES, 3'site:NO) | |
| XX | FT | 47939..49746 | |
| XX | FT | /tag= i | |
| XX | FT | /number= 5 | |
| XX | FT | /note= "Includes 17 base pairs of 5'UTR, the ORF and all of 3'UTR" | |
| XX | FT | 47939..47955 | |
| XX | FT | /tag= j | |
| XX | FT | /note= "Portion of 5' untranslated region (5'UTR)" | |
| XX | FT | 47956..49128 | |
| XX | FT | /tag= k | |
| XX | FT | /product= "Human glycosyl transferase-4alpha (GST-4alpha)" | |
| XX | FT | 49129..49746 | |
| XX | FT | /tag= l | |
| XX | FT | 83257..83347 | |
| XX | FT | /tag= m | |
| XX | FT | /label= 4a-5U2 | |
| XX | FT | 83348..96412 | |
| XX | FT | /tag= n | |
| XX | FT | /cons_splice= (5'site:NO, 3'site:NO) | |
| XX | FT | 96413..96484 | |
| XX | FT | /tag= o | |
| XX | FT | /label= 4a-5U1 | |
| XX | FT | 96485..98456 | |
| XX | FT | /tag= p | |
| XX | FT | /cons_splice= (5'site:NO, 3'site:NO) | |
| XX | FT | 98457..99968 | |
| XX | FT | /tag= q | |
| XX | FT | /note= "Includes 17 base pairs of 5'UTR, the ORF and all of 3'UTR" | |
| XX | FT | 98457..98473 | |
| XX | FT | /tag= r | |
| XX | FT | /note= "Portion of 5' untranslated region (5'UTR)" | |
| XX | FT | 98474..99661 | |
| XX | FT | /tag= s | |
| XX | FT | /product= "Human glycosyl transferase-4beta (GST-4beta)" | |
| XX | FT | 99662..99968 | |
| XX | FT | /tag= t | |
| XX | PD | WO200106015-A1. | |
| XX | PD | 25-JAN-2001. | |
| XX | PF | 19-JUL-2000; 2000WO-US19741. | |
| XX | PR | 20-JUL-1999; 99US-0144694. | |
| XX | PR | 13-JUL-2000; 2000US-0593828. | |
| XX | PA | (REGC) UNIV CALIFORNIA. | |
| XX | PI | Rosen SD, Lee JK, Hemmerich S; | |
| XX | DR | WPI: 2001-138471/14. | |
| XX | DR | P-PDB; AA72639, AA72640. | |
| XX | FT | New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications - | |

PS Example 1: Page 62-104; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
CC chromosome 16q23.1.

CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand. GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.

XX Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other.

SO Query Match 2.9%; Score 59; DB 22; Length 160552;
Best Local Similarity 100.0%; Pred. No. 2,5e-15;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 678 AGGCGTGGCTGCTACAGCCACGCTGCTGCTCAAGAGAGTGCGCTTCTCAACCTGCAG 736
|||||
Db 48446 AGGCTGCGCTCTACAGCCACGCTGCTGCTCAAGAGAGTGCGCTTCTCAACCTGCAG 48504

RESULT 19
AAZ20798
ID AAZ20798 standard; DNA; 37 BP.

XX AAZ20798;
AC AAZ20798;
XX 08-DEC-1999 (first entry)
DT
XX
XX PCR primer for glycosyl sulfotransferase-3 coding sequence.
DE
XX
XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; human;
KW secondary lymph organ; ss.
XX
XX Synthetic.
OS
XX
XX WC9945018-A1.
FN
XX
XX 30-SEP-1999.
PD
XX
XX 26-FEB-1999; 99WO-US04316.
PF
XX
XX 20-MAR-1998; 98US-0045284.
PR
XX
XX 12-NOV-1998; 98US-0190911.
PR
XX
XX (REGC) UNIV CALIFORNIA.
PA (SYNT) SYNTX USA INC.
PI
XX
XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX WPI; 1999-580442/49.
DR
XX
XX Human and murine glycosyl sulfotransferase 3 and related
PT polynucleotides -
XX
XX Example 4; Page 30; 59pp; English.
PS
XX
XX This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and

CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.

XX Sequence 37 BP; 14 A; 9 C; 8 G; 6 T; 0 other;

SO Query Match 1.8%; Score 37; DB 20; Length 37;
Best Local Similarity 100.0%; Pred. No. 9.9e-06;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 884 AAACTCAGAGAGGAGCCACACCTACTATGTGATGC 920
|||||
Db 1 AAACTCAGAGAGGAGGACCAACCTACTATGTGATGC 37

RESULT 20
ABK70821
ID ABK70821 standard; DNA; 30 BP.

XX ABK70821;
AC
XX
XX 15-JUL-2002 (first entry)
DT
XX
XX Probe for human gene CHST4.
DE
XX
XX Human: sulphuric acid conjugation; ss; probe; CHST1; CHST3;
KW CHST4; CHST5; CST; HNK-1ST; SULTAL; SULTBL; SULTX3; STE; TPST2.
KW
XX
XX Homo sapiens.
OS
XX
XX JP2002085067-A.
PN
XX
XX 26-MAR-2002.
PD
XX
XX 07-SEP-2000; 2000JP-0272229.
PF
XX
XX 07-SEP-2000; 2000JP-0272229.
PR
XX
XX (SANA) OTSUKA SEIRYAKU KOGYO KK.
PA
XX
XX WPI; 2002-378272/41.
DR
XX
XX Determination of enzymes participating in sulphuric acid conjugation in
PT humans, useful for confirmation of safety of investigational drugs.
PT
XX
XX comprises using oligonucleotide probes.
PT
XX
XX Claim 4; Page 9; 13pp; Japanese.
PS
XX
XX The invention relates to classification and quantitative determination of
CC enzymes participating in sulphuric acid conjugation comprising using
CC oligonucleotide probes hybridising to the following regions: (a) 885-911
CC region of CHST1 gene; (b) 174-197 region of CHST5 gene; (c) 1003-1032
CC region of CHST4 gene; (d) 322-346 region of CHST5 gene; (e) 737-765
CC region of CST gene; (f) 703-732 region of HNK-1ST gene; (g) 299-325
CC region of SUL2A1 gene; (h) 358-382 region of SUL2B1 gene; (i) 554-582
CC region of SULTX3 gene; (j) 451-478 region of STE gene; and (k) 652-677
CC region of TPST2 gene. Also included are PCR primers for the above
CC genes, kits and methods for determination. The probes, primers and the
CC method are used in the determination of sulphuric acid conjugation for
CC confirmation of the safety of investigational drugs. The present
CC sequence is an oligonucleotide probe for one of the above listed genes.
CC
XX
XX Sequence 30 BP; 5 A; 7 C; 7 G; 11 T; 0 other;

SO Query Match 1.5%; Score 30; DB 24; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0096;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1184 CGGTGCTTTGGCCATGAAAGGTTCT 1213
|||||
Db 1 CGGTGCTTTGGCCATGAAAGGTTCT 30

RESULT 21
AAZ20793
ID AAZ20793 standard; DNA; 1926 BP.
XX
AC AAZ20793;
XX
DT 08-DEC-1999 (first entry)
XX
DE Mouse glycosyl sulfotransferase-3 coding sequence.
XX
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ; ss.
XX
OS Mus sp.
XX
PN M09949018-A1.
XX
PD 30-SEP-1999.
XX
PE 26-FEB-1999; 99WO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
XX
PA (REGC) UNIV CALIFORNTA.
PA (SYNT) SYNTAX USA INC.
XX
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX
DR WPI: 1999-580442/49.
DR P-PSDB; AAY39919.
XX
PT Human and murine glycosyl sulfotransferase 3 and related
PT polynucleotides -
XX
PS Claim 4; Fig 3; 59pp; English.
XX
CC This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.
XX
SQ Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other;
Query Match 1.3%; Score 27; DB 20; Length 1926;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 494 GACATGACGCGCTTTGACGCTACATG 520
DB 714 GACATGACGCGCTTTGACGCTACATG 740
RESULT 22
AAC74310
ID AAC74310 standard; cDNA; 616 BP.
XX
AC AAC74310;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein gene 31 SEQ ID NO:41.
XX
KW Human: secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
KW cerebroprotective; neurotrophic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnary; gene therapy; angiogenesis;
DT

KW autoimmune disease; hyperproliferative disorder; infection; skin aging;
KW wound healing; cardiovascular disorder; cerebrovascular disorder;
KW nervous system disorder; food additive; preservative; ss.
XX
OS Homo sapiens.
XX
PN W0200057903-A2.
XX
PD 05-OCT-2000.
XX
PE 22-MAR-2000; 2000WO-US07525.
XX
PR 26-MAR-1999; 99US-0126595.
PR 22-DEC-1999; 99US-0171549.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI: 2000-584630/56.
DR P-PSDB; AAB39340.
XX
PT New nucleic acid molecules encoding 48 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; Page 339; 395pp; English.
XX
CC The polynucleotide sequences given in AAC74280 to AAC74327 encode the
CC human secreted proteins given in AAB39310 to AAB39357. AAB39358 to
CC AAB39400 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC antirheumatic; immunosuppressive; antiproliferative;
CC cytostatic; cardiac; vasotropic; cerebroprotective; neurotrophic;
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC and vulnary. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They can also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders, cardiovascular
CC disorders, cerebrovascular disorders, angiogenesis, nervous system
CC disorders, infections caused by bacteria, viruses and fungi and ocular
CC disorders. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors and other nutritional
CC components. AAC74271 to AAC74279 and AAB39309 represent sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 616 BP; 216 A; 102 C; 128 G; 169 T; 1 other;
Query Match 1.1%; Score 23; DB 21; Length 616;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2010 CAATTAAGAAAAA 2032
DB 570 CAATTAAGAAAAA 592
RESULT 23
AAS41008
ID AAS41008 standard; cDNA; 1142 BP.
XX
AC AAS41008;
XX
DT 17-DEC-2001 (first entry)

KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 8L1, ss.
 XX
 OS Mus musculus.
 XX
 XX Key Location/Qualifiers
 XX CDS 107..1294
 XX /tag= a
 XX /product= "Mouse glycosyl sulfortransferase-4 (GST-4)"
 XX /note= "CDS is specifically claimed as SEQ ID NO: 2
 XX in claim 6 (page no: 41) of the specification"
 XX
 XX WO200106015-A1.
 XX
 XX 25-JAN-2001.
 XX
 XX 19-JUL-2000; 2000WO-US19741.
 XX
 XX 20-JUL-1999; 9905-0144694.
 XX
 XX 13-JUL-2000; 2000US-0593828.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Rosen SD, Lee JK, Hemmerich S;
 XX WPI: 2001-138471/14.
 XX
 XX P-PSDS: AAY72638.
 XX
 XX New glycosyl sulfortransferases (GST)-alpha, GST-4beta and GST-6 for
 XX diagnostic and therapeutic agent screening applications -
 XX
 XX Claim 6; Fig 2; 128pp; English.
 XX
 XX The present sequence is mouse glycosyl sulfortransferase-4 (GST-4) cDNA.
 XX GST-4 gene is found on chromosome 8B1.
 XX GST-4 is a type 2 membrane protein useful for inhibiting a binding event
 XX between a selectin and a selectin ligand, which comprises contacting the
 XX selectin with a non-sulphated selectin ligand. GST and a small molecular
 XX agent that inhibits the sulphation activity of GST. GST is also useful
 XX in inhibiting a selectin mediated binding event. GST is useful in gene
 XX therapy to treat disorders such as acute or chronic inflammation,
 XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 XX disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, stress
 XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
 XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 XX during transplantation.
 XX
 XX Sequence 1989 BP; 415 A; 586 C; 550 G; 438 T; 0 other;
 XX
 XX Query Match 1.1%; Score 23; DB 22; Length 1989;
 XX Best Local Similarity 100.0%; Pred. No. 7.2;
 XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 704 GTGCTCAAGAGGTGGCTCTT 726
 XX |
 XX Db 620 GTGCTCAAGAGGTGGCTCTT 642
 XX
 XX RESULT 26
 XX ABV29163
 XX ID ABV29163 standard; cDNA; 2411 BP.
 XX
 XX AC ABV29163;
 XX
 XX

DT 16-SEP-2002 (first entry)
 XX
 XX Human prostate expression marker cDNA 29154.
 DE
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 KW
 XX Homo sapiens.
 OS
 XX
 XX WO200160860-A2.
 XX
 XX 23-AUG-2001.
 XX
 XX 20-FEB-2001; 2001WO-US05171.
 XX
 XX 17-FEB-2000; 2000US-183819P.
 XX
 XX 16-MAR-2000; 2000US-189862P.
 XX
 XX 25-MAY-2000; 2000US-207454P.
 XX
 XX 09-JUN-2000; 2000US-211314P.
 XX
 XX 18-JUL-2000; 2000US-219077P.
 XX
 XX 13-DEC-2000; 2000US-255281P.
 XX
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 XX Schlegel R, Endege WO, Monahan JE;
 XX WPI: 2001-662795/76.
 XX
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 XX Claim 1; Page 6203; 11750pp; English.
 XX
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 CC
 XX Sequence 2411 BP; 828 A; 476 C; 490 G; 617 T; 0 other;
 XX
 XX Query Match 1.1%; Score 23; DB 23; Length 2411;
 XX Best Local Similarity 100.0%; Pred. No. 7.1;
 XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 2010 CAATTAAGCAAAAAA 2032
 XX |
 XX Db 1999 CAATTAAGCAAAAAA 2021
 XX
 XX RESULT 27
 XX AAT63303/C
 XX ID AAT63303 standard; DNA; 2660 BP.
 XX
 XX AC AAT63303;
 XX
 XX 27-MAY-1997 (first entry)
 DE
 XX A. chrysogenum O-acetylhomoserine sulphydrylase genomic sequence.
 XX Acetmonium chrysogenum: O-acetylhomoserine sulphydrylase; chromatography;
 XX methionine synthase; PCR; polymerase chain reaction; primer; probe;
 KW

| | | |
|---|--|--|
| KM | | amplification; microorganism; cephalosporin; antibiotic; ds. |
| XX | | |
| OS | Acremonium chrysogenum. | |
| XX | | |
| Key | Location/Qualifiers | |
| FT CDS | /tag= a 449..2198 | |
| FT | /product= O-acetylhomoserine sulphydrylase | |
| FT exon | /tag= b 449..495 | |
| FT intron | /note= "exon 1" 496..646 | |
| FT | /tag= c /note= "intron 1" | |
| FT exon | 647..701 | |
| FT | /tag= d /note= "exon 2" | |
| FT intron | 702..771 | |
| FT | /tag= e /note= "intron 2" | |
| FT exon | 772..1215 | |
| FT | /tag= f /note= "exon 3" | |
| FT intron | 1216..1315 | |
| FT | /tag= g /note= "intron 3" | |
| FT exon | 1316..1914 | |
| FT | /tag= h /note= "exon 4" | |
| FT intron | 1915..1980 | |
| FT | /tag= i /note= "intron 4" | |
| FT exon | 1981..2044 | |
| FT | /tag= j /note= "exon 5" | |
| FT intron | 2045..2105 | |
| FT | /tag= k /note= "intron 5" | |
| FT exon | 2106..2195 | |
| FT | /tag= l /note= "exon 6" | |
| FN JP08336391-A. | | |
| XX | | |
| PD 24-DEC-1996. | | |
| XX PF 13-JUN-1995; | 95JP-0145866. | |
| PR 13-JUN-1995; | 95JP-0145866. | |
| PA (ASAH) ASAMI KASEI KOGYO KK. | | |
| XX MPI: 1997-102718/10. | | |
| DR P-PSDB; AAM12400. | | |
| PT O-acetylhomoserine sulphydrylase gene from Acremonium chrysogenum - | | |
| XX useful for producing the antibiotic cephalosporin C | | |
| PS Claim 4; Page 12-14; 15pp; Japanese. | | |
| XX | | |
| CC The sequence presented here is the genomic sequence of the Acremonium | | |
| CC chrysogenum O-acetylhomoserine sulphydrylase gene, more commonly known | | |
| CC as methionine synthase. The protein was isolated from A. chrysogenum | | |
| CC cells by conventional chromatographic methods and used for amino acid | | |
| CC sequencing. From peptide fragments of the protein, PCR primers | | |
| CC (AAT63304-5) were synthesised and used to amplify a fragment of the gene | | |
| CC for use as a probe to isolate the complete cDNA sequence (AAT63302). | | |
| CC The cDNA sequence was then used to isolate the genomic DNA sequence. | | |
| CC The gene can be used to transform microorganisms for the production of | | |
| CC cephalosporin C, a starting material for clinically important | | |
| XX cephalosporin type antibiotics. | | |
| Sequence 2860 BP; 570 A; 933 C; 766 G; 591 T; 0 other; | | |

```

Query Match      1.1%; Score 23; DB 18; Length 2860;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAGAAAAA 2032
|||||
Db 1248 CAAATAGAGAAAAA 1226

RESULT 28
ABLI6362/C
ID ABLI6362 standard; DNA: 4871 BP.
XX
XX ABLI6362:
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 559.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
CS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI: 2001-6556860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX
XX Claim 1; SEQ ID NO 559; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABBI7737-ABBI2072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 4871 BP: 1333 A; 1097 C; 1112 G; 1329 T; 0 other;

Query Match      1.1%; Score 23; DB 23; Length 4871;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAGAAAAA 2032
|||||
Db 674 CAAATAGAGAAAAA 652

RESULT 29
ABLI6364/C
ID ABLI6364 standard; DNA: 5101 BP.
XX

```

AC ABL16364;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 565.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 565; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL16175), expressed DNA
XX sequences (ABB57737-ABB72072),
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 5101 BP; 1420 A; 1129 C; 1153 G; 1399 T; 0 other;
SQ
Query Match 1.1%; Score 23; DB 23; Length 5101;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2010 CAAATTAAGAAAAA 2032
Db 904 CAAATTAAGAAAAA 882
RESULT 30
AB16360/C
ID ABL16360 standard; DNA; 10138 BP.
XX
XX ABL16360;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 553.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD

XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 553; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL16175), expressed DNA
XX sequences (ABB57737-ABB72072),
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 10138 BP; 3028 A; 2025 C; 2125 G; 2960 T; 0 other;
SQ
Query Match 1.1%; Score 23; DB 23; Length 10138;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2010 CAAATTAAGAAAAA 2032
Db 5941 CAAATTAAGAAAAA 5919
RESULT 31
ABK70834
ID ABK70834 standard; DNA; 22 BP.
XX
XX ABK70834;
XX
XX 15-JUL-2002 (first entry)
XX
XX PCR primer for human gene CHST4 #1.
XX
XX Human; sulphuric acid conjugation; ss: PCR; CHST1; CHST3; primer;
KW CHST4; CHST5; CST; HNK-1ST; SULT1A; SULT1B1; SULT1X3; STE; TPST2.
XX
XX Homo sapiens.
XX
XX JP2002085067-A.
XX
XX 26-MAR-2002.
XX
XX 07-SEP-2000; 2000JP-0272229.
XX
XX 07-SEP-2000; 2000JP-0272229.
XX
XX (SAKA) OTSUKA SEIYAKU KOGYO KK.
XX
XX WPI: 2002-378272/41.
XX
XX Determination of enzymes participating in sulphuric acid conjugation in
XX humans, useful for confirmation of safety of investigational drugs,
XX comprises using oligonucleotide probes -
XX
XX Claim 8; Page 10; 13pp; Japanese.

```
XX CC The invention relates to classification and quantitative determination of
CC enzymes participating in sulphuric acid conjugation comprising using
CC oligonucleotide probes hybridizing to the following regions: (a) 885-911
CC region of CHST1 gene; (b) 174-197 region of CHST3 gene; (c) 1003-1032
CC region of CHST4 gene; (d) 322-346 region of CHST5 gene; (e) 737-765
CC region of CST gene; (f) 703-732 region of HNK-1ST gene; (g) 299-325
CC region of SULTR2AI gene; (h) 358-382 region of SULTR1B1 gene; (i) 554-582
CC region of SULTR3 gene; (j) 451-478 region of STB gene; and (k) 652-677
CC region of TPST2 gene. Also included are PCR primers for the above
CC genes, kits and methods for determination. The probes, primers and the
CC method are used in the determination of sulphuric acid conjugation for
CC confirmation of the safety of investigational drugs. The present
CC sequence is a PCR primer for one of the above listed genes.
XX SQ Sequence 22 BP; 3 A; 8 C; 4 G; 7 T; 0 other:
XX
XX Query Match 1.1%; Score 22; DB 24; Length 22:
XX Best Local Similarity 100.0%; Pred. No. 25;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1161 CCCTATATGCTCCAGGCTTG 1182
Db 1 CCCTAATGCTCCAGGCTTG 22
XX
XX RESULT 32
XX AAS16960/C
XX ID AAS16960 standard; DNA; 22 BP.
XX AC AAS16960;
XX
XX 12-MAR-2002 (first entry)
XX
XX Human L-selectin sulfotransferase-2 (LST-2) cDNA RT-PCR primer #2.
DE
XX
XX Human: beta1,3gnt; beta1,3-N-acetylglicosaminyltransferase; MECA-79; ss;
XX L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
XX ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
XX allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
XX delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;
XX antiinflammatory; antiporiatic; antidiabetic; dermatological;
XX antiallergic; PCR primer; reverse transcriptase.
XX
XX Homo sapiens.
XX
XX WO200185177-A1.
XX
XX 15-NOV-2001.
XX
XX 10-MAY-2001; 2001WO-US15452.
XX
XX 11-MAY-2000; 2000US-0569320.
XX
XX (BURN-) BURNHAM INST.
XX
XX Fukuda M, Yeh J, Hiraoaka N;
XX
XX WPI; 2002-075226/10.
XX
XX New enzyme, useful for modifying acceptor molecule, comprises an
XX isolated L-selectin sulfotransferase-2 that directs expression of
XX L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
XX intestinal GlcNAc 6-sulfotransferase
XX
XX Example 2; Page 49; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor molecule
XX by contacting the acceptor with an isolated
XX beta1,3-N-acetylglicosaminyltransferase (beta1,3gnt) or an active
XX fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
XX invention also provides a method of treating or preventing an
XX L-selectin-mediated condition by reducing the expression or activity of a
```

```
CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
CC by administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3gnt,
CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1,3gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents a reverse transcriptase PCR (RT-PCR) primer used to
CC clone cDNA encoding human LST-2.
XX SQ Sequence 22 BP; 6 A; 5 C; 7 G; 4 T; 0 other:
XX
XX Query Match 1.1%; Score 22; DB 24; Length 22:
XX Best Local Similarity 100.0%; Pred. No. 25;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 538 ACAGTCCAGCCTTTCACTGG 559
Db 22 ACAGTCCAGCCTTTCACTGG 1
XX
XX RESULT 33
XX AAS16962/C
XX ID AAS16962 standard; DNA; 22 BP.
XX AC AAS16962;
XX
XX 12-MAR-2002 (first entry)
XX
XX Human L-selectin sulfotransferase-2 (LST-2) cDNA RT-PCR primer #4.
DE
XX
XX Human: beta1,3gnt; beta1,3-N-acetylglicosaminyltransferase; MECA-79; ss;
XX L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
XX ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
XX allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
XX delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;
XX antiinflammatory; antiporiatic; antidiabetic; dermatological;
XX antiallergic; PCR primer; reverse transcriptase.
XX
XX Homo sapiens.
XX
XX WO200185177-A1.
XX
XX 15-NOV-2001.
XX
XX 10-MAY-2001; 2001WO-US15452.
XX
XX 11-MAY-2000; 2000US-0569320.
XX
XX (BURN-) BURNHAM INST.
XX
XX Fukuda M, Yeh J, Hiraoaka N;
XX
XX WPI; 2002-075226/10.
XX
XX New enzyme, useful for modifying acceptor molecule, comprises an
XX isolated L-selectin sulfotransferase-2 that directs expression of
XX L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
XX intestinal GlcNAc 6-sulfotransferase
XX
XX Example 2; Page 49; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor molecule
XX by contacting the acceptor with an isolated
XX beta1,3-N-acetylglicosaminyltransferase (beta1,3gnt) or an active
XX fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
XX invention also provides a method of treating or preventing an
```

CC L-selectin-mediated condition by reducing the expression or activity of a
 CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3gnt,
 CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LSSST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LSSST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis.
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents a reverse transcriptase PCR (RT-PCR) primer used to
 CC clone cDNA encoding human LSSST-2.
 XX
 SQ Sequence 22 BP: 5 A; 6 C; 5 G; 6 T; 0 other:
 Query Match 1.1%; Score 22; DB 24; Length 22;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 914 GTGATGAGCTCATCTGCCAA 935
 Db 22 GTGATGAGCTCATCTGCCAA 1
 RESULT 34
 AAS16964/c
 ID AAS16964 standard; DNA: 22 BP.
 XX
 AC AAS16964:
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Human L-selectin sulfotransferase-2 (LSSST-2) cDNA RT-PCR primer #6.
 XX
 KW Human: beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen Planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSSST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antitumor;
 KW antiinflammatory; antiproliferative; antidiabetic; dermatological;
 KW antiallergic; PCR primer; reverse transcriptase.
 XX
 OS Homo sapiens.
 OS
 PN WO200185177-A1.
 XX
 PD 15-NOV-2001.
 XX
 PE 10-MAY-2001; 2001WO-US16452.
 XX
 PR 11-MAY-2000; 2000US-0569920.
 XX
 PA (BURN-) BURNHAM INST.
 XX
 PI Fukuda M, Yeh J, Hiraoaka N;
 XX
 DR WPI; 2002-075226/10.
 XX
 PT New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 PT intestinal GlcNAc 6-sulfotransferase
 XX
 PS Example 2; Page 49; 98pp; English.
 XX
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active
 CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The

CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3gnt,
 CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LSSST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LSSST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents a reverse transcriptase PCR (RT-PCR) primer used to
 CC clone cDNA encoding human LSSST-2.
 XX
 SQ Sequence 22 BP: 8 A; 3 C; 8 G; 3 T; 0 other:
 Query Match 1.1%; Score 22; DB 24; Length 22;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1401 CTTCTGAGCTTGCCTACATCT 1422
 Db 22 CTTCTGAGCTTGCCTACATCT 1
 RESULT 35
 ABV38642/c
 ID ABV38642 standard; cDNA: 341 BP.
 XX
 AC ABV38642:
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 38633.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PE 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer
 XX
 PS Claim 1; Page 7862; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 341 BP; 81 A; 76 C; 83 G; 101 T; 0 other;
Query Match 1.1%; Score 22; DB 23; Length 341;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2011 AATAGAGAAAAA 2032
D 85 AATAGAGAAAAA 64
RESULT 36
ABV54232 standard; cDNA: 502 BP.
ID ABV54232 standard; cDNA: 502 BP.
XX
AC ABV54232;
XX
DT 17-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 54223.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207434P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI: 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 10489; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 502 BP; 227 A; 87 C; 98 G; 90 T; 0 other;
Query Match 1.1%; Score 22; DB 23; Length 502;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2011 AATAGAGAAAAA 2032
D 137 AATAGAGAAAAA 158
RESULT 37
AAS41556 standard; cDNA: 1781 BP.
ID AAS41556 standard; cDNA: 1781 BP.
XX
AC AAS41556;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #772.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 28-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220863.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225470.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0235837.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251865.
PR 08-DEC-2000; 2000US-0251985.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465566/50.
XX P-PSDB; AAU23686.
DR
XX
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
XX
PS Claim 4; SEQ ID NO 782; 1180pp; English.
XX
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU2915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU40785-AAU41684 represent cDNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 1781 BP; 535 A; 331 C; 395 G; 518 T; 2 other;

Query Match 1.1%; Score 22; DB 22; Length 1781;
BestLocal Similarity 100.0%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2011 AATTAAGAAAAA 2032
 |||||||
 Db 1758 AATTAAGAAAAA 1779
 |||||||

RESULT 38
 ABR63555
 ID ABR63555 standard; cDNA: 1782 BP.
 XX
 AC ABR63555;
 XX
 DI 18-JUN-2002 (first entry)
 XX
 DE Rat sequence differentially expressed in response to a hepatotoxin #1462.
 XX
 KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 KW differential expression; centrilobular necrosis; steatosis.
 XX
 OS Rattus norvegicus.
 XX
 PN WO200210453-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 30-JUL-2001; 2001WO-US23872.
 XX
 PR 31-JUL-2000; 2000US-222040P.
 PR 02-NOV-2000; 2000US-244880P.
 PR 11-MAY-2001; 2001US-290029P.
 PR 15-MAY-2001; 2001US-290645P.
 PR 22-MAY-2001; 2001US-292336P.
 PR 06-JUN-2001; 2001US-295798P.
 PR 13-JUN-2001; 2001US-297457P.
 PR 19-JUN-2001; 2001US-298884P.
 PR 09-JUL-2001; 2001US-303459P.
 XX
 XX (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff NR;
 XX
 DR WPI: 2002-241625/29.
 XX
 XX Predicting toxic effects of compounds or the progression of these toxic
 PT effects by determining the changes in gene expression in tissues or
 PT cells exposed to the toxin and comparing these to gene expression in
 PT unexposed tissues or cells -
 XX
 PS Claim 1; Seq ID No 1462; 239pp; English.
 XX
 CC The invention relates to methods for predicting toxic effects of
 CC compounds or the progression of these toxic effects by determining the
 CC global changes in gene expression in tissues or cells exposed to the
 CC toxin and comparing these to gene expression in unexposed tissues or
 CC cells. Also included are methods of predicting at least one toxic
 CC effect of a compound or progression of a toxic effect, preferably the
 CC hepatotoxicity of a compound, comprising detecting the level of
 CC expression in a tissue or cell sample exposed to the compound of two or
 CC more genes listed in the specification, where differential expression of
 CC the genes is indicative of at least one toxic effect or progression.
 CC The method can also be used to identify an agent which modulates the
 CC toxic response and predict cellular pathways that a compound modulates
 CC in a cell. The methods utilise a set of at least two probes (on a solid
 CC support in kit form), where each of the probes comprises a sequence that
 CC specifically hybridises to a gene listed in the specification, a computer
 CC system comprising a database containing information identifying the
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
 CC set of genes comprising at least two genes listed in the specification,
 CC and a user interface to view the information used to present information
 CC identifying the expression level in a tissue or cell of at least one gene
 CC listed in the specification. The method is useful for elucidating global
 CC changes in gene expression and for identifying toxicity markers in
 CC tissues or cell exposed to a known toxin. The genes may be used as
 CC toxicity markers in drug screening and toxicity assays. The genes and

CC gene expression information may be used as diagnostic markers for the
 CC prediction or identification of the physiological state of tissue or cell
 CC sample that has been exposed to a compound or agent. Hepatotoxicity
 CC is characterised by centrilobular necrosis and steatosis. The present
 CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
 CC which is differentially expressed in response to a hepatotoxic agent.
 XX
 SQ Sequence 1782 BP; 413 A; 516 C; 474 G; 379 T; 0 other;
 XX

Query Match 1.1%; Score 22; DB 24; Length 1782;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2011 AATTAAGAAAAA 2032
 |||||||
 Db 1756 AATTAAGAAAAA 1777
 |||||||

RESULT 39
 AAH75633/C
 ID AAH75633 standard; cDNA: 2415 BP.
 XX
 AC AAH75633;
 XX
 DI 31-OCT-2001 (first entry)
 XX
 DE Human ribosomal protein S4-19 encoding cDNA.
 XX
 KW Human; ribosomal protein S4-19; cytosolic; virucidal; immunomodulatory;
 KW antiinflammatory; haemostatic; anti-HIV; malignant tumour; HIV;
 KW human immunodeficiency virus; infection; immunological disease;
 KW gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1817..2356
 FT CDS /tag= a
 FT /product= "ribosomal protein S4-19"
 FT /note= "Claimed in claim 6"
 XX
 PN WO200166581-A1.
 XX
 PD 13-SEP-2001.
 XX
 PE 26-FEB-2001; 2001WO-CN00162.
 XX
 PR 07-MAR-2000; 2000CN-0111943.
 XX
 PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
 XX
 PI Mao Y, Xie Y;
 XX
 DR WPI: 2001-565568/63.
 DR P-PSDB; AAG77991.
 XX
 PT New human ribosomal protein S4-19 for diagnosing and treating malignant
 PT tumour, haemopathy, human immunodeficiency virus infection, immunological
 PT diseases and inflammations -
 XX
 PS Claim 6; Page 29-30; 35pp; Chinese.
 XX
 CC The invention relates to the human ribosomal protein S4-19 with
 CC cytosolic, virucidal, immunomodulatory, antiinflammatory, haemostatic
 CC and anti-HIV. The polypeptide and encoded polynucleotide are applicable
 CC in diagnosis and treatment of malignant tumour, haemopathy, human
 CC immunodeficiency virus (HIV) infection, immunological diseases and
 CC various inflammation. The polynucleotide is useful for gene therapy. The
 CC present sequence is that of the human ribosomal protein S4-19 encoding
 CC cDNA.
 XX
 SQ Sequence 2415 BP; 611 A; 530 C; 599 G; 675 T; 0 other;

```
Query Match 1.1%; Score 22; DB 22; Length 2415;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2011 AATAGAGAAAAA 2032
Db 1199 AATAGAGAAAAA 1178

RESULT 40
AAQ11712/c
ID AAQ11712 standard; DNA: 3138 BP.
XX
XX AAQ11712;
AC
XX 30-JUL-1991 (first entry)
DT
XX
XX Shuttle vector PMCM1630.
DE
XX
XX slime mould; replication; Rep gene; Ddp2; ss.
XX
XX Dictyostelium discoideum.
OS
XX
XX Key Location/Qualifiers
FH sig_peptide 2068
FT /tag= a
FT /note= "finish position not given"
XX
XX W09106644-A.
XX
XX 16-MAY-1991.
PD
XX
XX 02-NOV-1980; 90WO-AU00530.
PE
XX
XX 02-NOV-1989; 89AU-0007187.
PR
XX
XX (UYMA-) MACQUARIE UNIV.
PA
XX
XX Slade MB, Chang ACM, Williams KL;
PI
XX
XX WPI: 1991-164194/22.
DR
XX
XX Polypeptide facilitating extra-chromosomal replication - of
PT recombinant plasmid in Dictyostelium species
PS
XX
XX Claim 35; Fig 15; 90pp; English.
PS
XX
XX Shuttle vector PMCM1630 was constructed for use in E.coli and
CC transfer back into Dictyostelium. It includes a fragment of the
CC Dictyostelium plasmid Ddp2 containing the origin of replication,
CC the promoter from the Dictyostelium Actin 15 gene, a sequence
CC encoding the secretion signal peptide of the D19 gene of protein
CC Psa, Actin 15 polyA signal and an ampicilline resistance-conferring
CC sequence. See also AAQ11710 and AAQ11711.
CC
XX
XX Sequence 3138 BP; 979 A; 568 C; 574 G; 1017 T; 0 other;
SO

Query Match 1.1%; Score 22; DB 12; Length 3138;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2011 AATAGAGAAAAA 2032
Db 2030 AATAGAGAAAAA 2009

RESULT 41
ABA82684
ID ABA82684 standard; DNA: 3597 BP.
XX
XX ABA82684;
AC
XX
XX 25-JAN-2002 (first entry)
DT
```

```
XX
XX HSM800936 gene SEQ ID NO:70.
DE
XX
XX Human: high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW sentence tagged site; SRS; osteoporosis; osteopathic; gene therapy;
KW antinease therapy; vaccine; bone disorder; Paget's disease;
KW sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
XX Homo sapiens.
OS
XX
XX W0200177327-A1.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 21-JUN-2000; 2000WO-US16951.
PE
XX
XX 05-APR-2000; 2000US-0543771.
PR
XX
XX 05-APR-2000; 2000US-0544398.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Carulli JP, Little RD, Recker RR, Johnson ML;
PI
XX
XX WPI: 2001-657171/75.
DR
XX
XX New high bone mass (HBM) and Zmax1 genes and proteins useful for
PT modulating bone mass for the treatment of e.g. osteoporosis -
XX
XX Claim 79; Page 372-373; 443pp; English.
PS
XX
XX The present invention describes the human Zmax1 gene and the high bone
CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
CC HBM genes have osteopathic activities. The genes can be used in gene
CC therapy, antinease therapy and in the production of vaccines. They
CC can be used in the diagnosis and treatment of bone disorders including
CC osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
CC dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
CC sequences used in the exemplification of the present invention.
CC
XX
XX Sequence 3597 BP; 981 A; 929 C; 776 G; 911 T; 0 other;
SO

Query Match 1.1%; Score 22; DB 22; Length 3597;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2011 AATAGAGAAAAA 2032
Db 3571 AATAGAGAAAAA 3592

RESULT 42
ABLO7230/c
ID ABLO7230 standard; CDNA: 73947 BP.
XX
XX ABL07230;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 16172.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX W0200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PE
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PR
```

XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li FMD, Myers EM,
PI WPI: 2001-656860/75.
XX P-PSDB: ABB63127.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 16172; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB160511), expressed DNA
CC sequences (AB161840-AB161175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_ptl_sequences.
XX
SO Sequence 73947 BP; 21410 A; 14389 C; 15751 G; 22397 T; 0 other:

Query Match 1.1%; Score 22; DB 23; Length 73947;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATTAAGAAAAA 2031
DB 73335 CAATTAAGAAAAA 73314

RESULT 43
ABK70835/c
ID ABK70835 standard; DNA: 21 BP.
XX
XX ABE70835;
XX
XX 15-JUL-2002 (first entry)
XX
XX PCR primer for human gene CHST4 #2.
XX
XX Human: sulphuric acid conjugation; ss; PCR: CHST1; CHST3; primer;
KM CHST4; CHST3; CST; HNK-1ST; SULT1A1; SULT1X3; STE; TPST2.
XX
XX Homo sapiens.
XX
XX OS
XX PN JP2002085067-A.
XX
XX PD 26-MAR-2002.
XX
XX 07-SEP-2000; 2000JP-0272229.
XX
XX 07-SEP-2000; 2000JP-0272229.
XX
XX (SAKA) OTSUKA SEIYAKU KOGYO KK.
XX
XX WPI: 2002-378272/41.
XX
XX Determination of enzymes participating in sulphuric acid conjugation in
PT humans, useful for confirmation of safety of investigational drugs,
PT comprises using oligonucleotide probes -
XX
XX Claim 8; Page 11; 13pp; Japanese.
XX
XX The invention relates to classification and quantitative determination of
CC enzymes participating in sulphuric acid conjugation comprising using
CC oligonucleotide probes hybridising to the following regions: (a) 885-911

CC region of CHST1 gene; (b) 174-197 region of CHST3 gene; (c) 1003-1032
CC region of CHST4 gene; (d) 322-346 region of CHST5 gene; (e) 737-765
CC region of CST gene; (f) 703-732 region of HNK-1ST gene; (g) 299-325
CC region of SULT2A1 gene; (h) 358-382 region of SULT2B1 gene; (i) 554-582
CC region of SULT3 gene; (j) 451-478 region of STE gene; and (k) 652-677
CC region of TPST2 gene. Also included are PCR primers for the above
CC genes, kits and methods for determination. The probes, primers and the
CC method are used in the determination of sulphuric acid conjugation for
CC confirmation of the safety of investigational drugs. The present
CC sequence is a PCR primer for one of the above listed genes.
XX
SO Sequence 21 BP; 6 A; 6 C; 5 G; 4 T; 0 other:

Query Match 1.0%; Score 21; DB 24; Length 21;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1240 CATGAATTGCTGGGCTACCG 1260
DB 21 CATGAATTGCTGGGCTACCG 1

RESULT 44
AAS16961
ID AAS16961 standard; DNA: 21 BP.
XX
XX AAS16961;
XX
XX 12-MAR-2002 (first entry)
XX
XX Human L-selectin sulfotransferase-2 (LST-2) cDNA Rr-PCR primer #3.
XX
XX
XX Human: beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
KM L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KM ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
KM allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
KM delayed-type hypersensitivity reaction; hyperplastic thymus; antileuk;
KM antiinflammatory; antiproliferative; antidiabetic; dermatological;
KM antiallergic; PCR primer: reverse transcriptase.
XX
XX OS
XX PN WO200185177-A1.
XX
XX PD 15-NOV-2001.
XX
XX 10-MAY-2001; 2001WO-US15452.
XX
XX 11-MAY-2000; 2000US-0569320.
XX
XX (BURN-) BURHAM INST.
XX
XX Fukuda M, Yeh J, Hiraoaka N;
XX
XX WPI: 2002-075226/10.
XX
XX New enzyme, useful for modifying acceptor molecule, comprises an
PT isolated L-selectin sulfotransferase-2 that directs expression of
PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
PT intestinal GlcNAc 6-sulfotransferase -
XX
XX Example 2; Page 49; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated
CC beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active
CC fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The
CC invention also provides a method of treating or preventing an
CC L-selectin-mediated condition by reducing the expression or activity of a
CC beta1,3GNT that directs expression of a MECA-79 antigen. This can be done
CC by administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3GNT,

CC and/or a beta1.3gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1.3gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents a reverse transcriptase PCR (RT-PCR) primer used to
CC clone cDNA encoding human LST-2.

XX Sequence 21 BP; 0 A; 7 C; 6 G; 8 T; 0 other;

Query Match 1.0%; Score 21; DB 24; Length 21;
Best Local Similarity 100.0%; Pred. No. 67;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 319 GGTCCTGCTCTCTGGCGCTC 339
|||||

Db 1 GGTCCTGCTCTCTGGCGCTC 21

RESULT 45

AL27808/c
ID AL27808 standard; DNA: 51 BP.

XX AC AL27808;

XX DT 24-JAN-2002 (first entry)

XX DE Human SNP oligonucleotide #1016.

XX Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
XX neuroprotective; antimicrobial; gene therapy; vaccine; amyase; cancer;
XX amyloid protein; angiotensin; apoptosis related protein; cadherin;
XX cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
XX complement related protein; cytochrome; kinesin; cytokine; interferon;
XX interleukin; G-protein coupled receptor; thioesterase; inflammation;
XX multifactorial disease; autoimmune disease; infection;
XX nervous system disease; ss.

XX OS Homo sapiens.

XX PN WC200147944-A2.

XX PD 05-JUL-2001.

XX PF 28-DEC-2000; 2000WC-US35498.

XX PR 28-DEC-1999; 9905-0173419.

XX PR 27-DEC-2000; 2000US-0173419.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach M;

XX DR WPI: 2001-465210/50.

XX PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
XX PT oncogenes and histones, useful for diagnosing and treating, e.g.

XX PT cancer, autoimmune diseases and infections -

XX PS Claim 1; Page 1670; 4143pp; English.

XX The present invention relates to oligonucleotides encoding polymorphic
CC variants of proteins related to amylases, amyloid proteins, angiotensin,
CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
CC histones, kinases, colony stimulating factors, complement related
CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,
CC G-protein coupled receptors and thioesterases. The present sequence is
CC one such oligonucleotide. The oligonucleotides and the peptides encoded
CC by them may be used in the prevention, diagnosis and treatment of

CC diseases associated with inappropriate expression of the proteins listed
CC above. Disorders that may be prevented, diagnosed and/or treated include
CC multifactorial diseases with a genetic component, such as autoimmune
CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
CC systemic lupus erythematosus and Grave's disease), inflammation, cancer
CC (e.g. cancers of the bladder, brain, breast, colon and kidney,
CC leukemia), diseases of the nervous system and an infection of pathogenic
CC organisms.

XX Sequence 51 BP; 6 A; 4 C; 3 G; 38 T; 0 other;

Query Match 1.0%; Score 21; DB 22; Length 51;
Best Local Similarity 100.0%; Pred. No. 63;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2012 AATACAAAAA 2032
|||||

Db 29 AATACAAAAA 9

RESULT 46

ABV08745/c
ID ABV08745 standard; CDNA: 262 BP.

XX AC ABV08745;

XX DT 13-SEP-2002 (first entry)

XX DE Human prostate expression marker CDNA 8736.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200150860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US05171.

XX PR 17-FEB-2000; 2000US-18319P.

XX PR 16-MAR-2000; 2000US-189662P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

XX PR 13-DEC-2000; 2000US-255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX DR WPI: 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer -

XX PS Claim 1; Page 1378-1379; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient,
CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 262 BP; 56 A; 56 C; 58 G; 88 T; 4 other;

Query Match 1.0%; Score 21; DB 23; Length 262;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

CY 2012 AATAAGAAAAAAAAAAAAA 2032
 |||||
DB 33 AATAAGAAAAAAAAAAAAA 13

RESULT 47
AAF63356
ID AAF63356 standard; DNA; 294 BP.
XX
AC AAF63356;
XX
DT 04-MAY-2001 (first entry)
XX
DE DNA encoding larval T. solium TSRS1 protein.
XX
KW Larval Taenia solium; cysticercosis; pork tapeworm; vaccine; TSRS1; ds.
XX
OS Taenia solium.
XX
Key Location/Qualifiers
FH 3, 227
FT /*tag= a
FT /product= "TSRS1"
FT /partial
XX
FN WO200110897-A2.
PD 15-FEB-2001.
XX
PF 03-AUG-2000; 2000WO-US21173.
XX
PR 05-AUG-1999; 99US-0147318.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Tsang VCM, Greene RM, Wilkins PP, Hancock K;
PL WPI: 2001-202757/20.
DR P-PADB; AAB72214.
XX
PT Composition for detecting larval Taenia solium, comprising isolated,
PT synthetic or recombinant larval Taenia solium polypeptides that are
PT immunoreactive with Taenia solium antibodies .
XX
XX
XX Claim 7; Page 34-35; 37pp; English.

XX This invention relates to a composition comprising one or more isolated,
XX synthetic or recombinant larval Taenia solium polypeptides, or its
XX antigenic fragments. The invention includes T. solium polypeptides TS-14,
XX TS-18 and TSRS-1 and their encoding polynucleotide sequences, also
XX included are T. solium antigenic peptides. A T. solium polypeptide is
XX useful for diagnosing cysticercosis in a mammal. T. solium is also
XX useful for detecting T. solium antibodies in a biological sample. The
XX proteins are useful for reducing, possibly preventing, T. solium
XX infection or transmission and are used in immunoassays for detecting
XX T. solium. Nucleic acid sequences encoding the proteins are useful as
XX molecular probes or primers for detecting RNA and DNA involved
XX transcription and translation of the proteins. The proteins and
XX polynucleotide sequences may be used to form a vaccine which will protect
XX against T. solium also known as pork tapeworm. The present sequence
XX represents DNA encoding TSRS1.

SQ Sequence 294 BP; 107 A; 50 C; 71 G; 66 T; 0 other;

Query Match 1.0%; Score 21; DB 22; Length 294;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

CY 2012 AATAGAAAAAAAAAAAAA 2032
| | | | |
DB 271 AATAGAAAAAAAAAAAAA 291

RESULT 48
ABV56921/C
ID ABV56921 standard; CDNA; 397 BP.
XX
XX AC ABV56921;
XX DT 17-SEP-2002 (first entry)
XX DE
XX EE Human prostate expression marker CDNA 56912.
XX DE
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KM pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX SS
XX PN WO200160860-A2.
XX PD
XX PF 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US05171.
XX PR
XX PR 17-FEB-2000; 2000US-18319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX
XX PI Schlegel R, Endege WO, Monahan JE;
PI
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer; useful
PT for detecting presence of prostate cancer; stage of prostate cancer -
XX
PS Claim 1; Page 10964; 11750bp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213), of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 397 BP; 75 A; 89 C; 87 G; 146 T; 0 other;

Query Match 1.0%; Score 21; DB 23; Length 397;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 2012 AATAGAAAAAAAAAAAAA 2032

Mon Jan 13 09:13:58 2003

us-09-816-825-1.ol120.rst

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 21:44:02 ; Search time 2885 Seconds

(without alignments)
11407.012 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032

Sequence: 1 ggctcagagccagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO_NUC
Gapox 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 20

Total number of hits satisfying chosen parameters: 1660

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

EST:*
1: em_estba:*
2: em_esthm:*
3: em_estim:*
4: em_estmu:*
5: em_estoy:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pio:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

| SUMMARIES | | | |
|-----------|-------|--------------------|---------------------------|
| Result. | % | Query | Description |
| No. | Score | Match length DB ID | |
| 1 | 630 | 31.0 | 954 13 B1823850 603039012 |
| 2 | 581 | 28.6 | 593 14 BM969292 |
| 3 | 551 | 27.1 | 571 13 BM129080 |
| 4 | 478 | 23.5 | 553 13 BM128370 |
| 5 | 478 | 23.5 | 553 13 BM128370 |
| 6 | 468 | 23.0 | 668 9 AL709927 |

| | | | | |
|----|-----|------|------------------|---------------------|
| 7 | 318 | 15.6 | 680 12 BE878439 | BE878439 MR0-ET010 |
| 8 | 288 | 14.2 | 362 10 AM002418 | AM002418 wu61g03.x |
| 9 | 283 | 13.9 | 419 10 AM572390 | AM572390 x09a06.x |
| 10 | 150 | 7.4 | 417 10 BM572510 | BM572510 x918g11.x |
| 11 | 136 | 6.7 | 358 12 BE056840 | BE056840 7x10d12.x |
| 12 | 59 | 2.9 | 525 12 BF197521 | BF197521 7084d08.x |
| 13 | 59 | 2.9 | 620 9 A1824100 | A1824100 w146c01.x |
| 14 | 59 | 2.9 | 695 12 BE857538 | BE857538 7901a08.x |
| 15 | 59 | 2.9 | 635 10 BE858652 | BE858652 7901a09.x |
| 16 | 47 | 2.3 | 436 12 AM081348 | AM081348 xc41b06.x |
| 17 | 27 | 1.3 | 497 13 BM246681 | BM246681 k0741804- |
| 18 | 27 | 1.3 | 536 13 BM245312 | BM245312 .k072207- |
| 19 | 27 | 1.3 | 695 13 BG364671 | BG364671 602831875 |
| 20 | 27 | 1.3 | 783 13 BG363298 | BG363298 602827716 |
| 21 | 27 | 1.3 | 852 13 BG366340 | BG366340 602832826 |
| 22 | 27 | 1.3 | 1923 11 AK009113 | AK009113 Mus.muscu |
| 23 | 25 | 1.2 | 501 13 BM036473 | BM036473 fu81f07.x |
| 24 | 25 | 1.2 | 682 17 CNO02BKQ | AL189971 Tetraodon |
| 25 | 24 | 1.2 | 375 13 BM663231 | BM663231 UI-E-DX1- |
| 26 | 24 | 1.2 | 566 13 BM661130 | BM661130 UI-E-DX1- |
| 27 | 24 | 1.2 | 919 17 CNO01L7X | AL173359 Tetraodon |
| 28 | 24 | 1.1 | 132 17 B75634 | B75634 PC111-1IM1 |
| 29 | 23 | 1.1 | 191 9 A1072478 | A1072478 UI-R-C2-n |
| 30 | 23 | 1.1 | 269 10 B128285 | B128285 B128285 |
| 31 | 23 | 1.1 | 304 14 BQ191911 | BQ191911 UI-R-DRI- |
| 32 | 23 | 1.1 | 319 10 BE117198 | BE117198 UI-R-B51- |
| 33 | 23 | 1.1 | 342 12 BG291580 | BG291580 6028185671 |
| 34 | 23 | 1.1 | 349 13 B1555949 | B1555949 603339076 |
| 35 | 23 | 1.1 | 362 10 BE340900 | BE340900 EST344988 |
| 36 | 23 | 1.1 | 385 14 BM930440 | BM930440 UI-E-ET1- |
| 37 | 23 | 1.1 | 397 13 B1505360 | B1505360 B170026A |
| 38 | 23 | 1.1 | 408 10 BE109147 | BE109147 UI-R-B51- |
| 39 | 23 | 1.1 | 413 9 A1095726 | A1095726 qb24f12.x |
| 40 | 23 | 1.1 | 419 10 AM080577 | AM080577 xc3se07.x |
| 41 | 23 | 1.1 | 420 10 BE107411 | BE107411 UI-R-B51- |
| 42 | 23 | 1.1 | 423 9 A1195993 | A1195993 u163b09.x |
| 43 | 23 | 1.1 | 437 14 W32300 | W32300 zc05q05.s1 |
| 44 | 23 | 1.1 | 443 9 A1952475 | A1952475 wx274604.x |
| 45 | 23 | 1.1 | 446 14 BQ013210 | BQ013210 UI-L-BCIP |
| 46 | 23 | 1.1 | 461 14 BQ427521 | BQ427521 AGENCOURT |
| 47 | 23 | 1.1 | 481 14 BQ194968 | BQ194968 UI-R-CM1- |
| 48 | 23 | 1.1 | 484 10 AM434101 | AM434101 UI-R-BJOP |
| 49 | 23 | 1.1 | 484 12 BE062672 | BE062672 7h63g02.x |
| 50 | 23 | 1.1 | 497 9 AA809972 | AA809972 ca95f12.s |
| 51 | 23 | 1.1 | 498 13 BM316129 | BM316129 fw63g11.x |
| 52 | 23 | 1.1 | 501 9 AA765670 | AA765670 ca04f06.s |
| 53 | 23 | 1.1 | 538 10 BE419040 | BE419040 WMR018.C5 |
| 54 | 23 | 1.1 | 545 13 B1407057 | B1407057 189A02.Ma |
| 55 | 23 | 1.1 | 577 12 BG897008 | BG897008 HOA59-1-H |
| 56 | 23 | 1.1 | 616 17 A2424759 | A2424759 1M02A0G04 |
| 57 | 23 | 1.1 | 616 17 BE109266 | BE109266 UI-R-B51- |
| 58 | 23 | 1.1 | 626 17 AC183235 | AC183235 Pan.tl091 |
| 59 | 23 | 1.1 | 720 14 BQ772892 | BQ772892 UI-R-EEO- |
| 60 | 23 | 1.1 | 721 14 BQ775707 | BQ775707 UI-R-EHO- |
| 61 | 23 | 1.1 | 819 12 BE875409 | BE875409 601489128 |
| 62 | 23 | 1.1 | 834 14 BQ884877 | BQ884877 AGENCOURT |
| 63 | 23 | 1.1 | 857 13 B1517608 | B1517608 603042215 |
| 64 | 23 | 1.1 | 1048 13 BM393197 | BM393197 500712-7 |
| 65 | 23 | 1.1 | 1048 13 BM393024 | BM393024 50072-7 |
| 66 | 22 | 1.1 | 86 17 CNO02R89 | AL210258 Tetraodon |
| 67 | 22 | 1.1 | 67 14 BQ395065 | BQ395065 NISC-ng12 |
| 68 | 22 | 1.1 | 77 9 A1312401 | A1312401 ta80e08.x |
| 69 | 22 | 1.1 | 86 10 AA409770 | AA409770 fh02c07.y |
| 70 | 22 | 1.1 | 86 13 BM055842 | BM055842 1g83f09.y |
| 71 | 22 | 1.1 | 93 10 AV742926 | AV742926 AV742926 |
| 72 | 22 | 1.1 | 96 13 B1697249 | B1697249 603346757 |
| 73 | 22 | 1.1 | 135 13 B1404524 | B1404524 MT-E-CP1- |
| 74 | 22 | 1.1 | 139 17 A2896024 | A2896024 RPCT-24-1 |
| 75 | 22 | 1.1 | 152 13 B1268683 | B1268683 NF018B10G |
| 76 | 22 | 1.1 | 154 13 B1042721 | B1042721 CM4-OT016 |
| 77 | 22 | 1.1 | 155 10 BE655780 | BE655780 UI-W-AJ1- |
| 78 | 22 | 1.1 | 155 13 BM187070 | BM187070 fv81f11.y |
| 79 | 22 | 1.1 | 156 9 AU037435 | AU037435 AU037435 |

| | | | | | | | | | | | | | |
|-------|----|-----|-----|----|-----------|------------|-------|-----|------|----|-----------|-----------|-------------|
| C 226 | 22 | 1.1 | 583 | 12 | BG083901 | H3093D10- | 299 | 1.1 | 782 | 13 | B1661882 | B1661882 | 603305622 |
| C 227 | 22 | 1.1 | 584 | 13 | B1988624 | A013-41.M | C 300 | 1.1 | 794 | 9 | AA814807 | AA814807 | 0b73b05.s |
| C 228 | 22 | 1.1 | 586 | 14 | BM676242 | UT-E-EJO- | C 301 | 1.1 | 811 | 14 | BQ796616 | BQ796616 | EST.5554 |
| C 229 | 22 | 1.1 | 594 | 10 | BM651147 | UT-E-CII- | C 302 | 1.1 | 812 | 10 | BE541102 | BE541102 | 601063642 |
| C 230 | 22 | 1.1 | 600 | 10 | AV386013 | AV386013 | C 303 | 1.1 | 815 | 13 | BM360852 | BM360852 | GA_Ea003 |
| C 231 | 22 | 1.1 | 600 | 13 | B1711215 | 1d93d04.x | C 304 | 1.1 | 850 | 17 | AZ540576 | AZ540576 | ENTID627E |
| C 232 | 22 | 1.1 | 600 | 13 | BM310222 | 1h09e10.x | C 305 | 1.1 | 856 | 17 | CNS0530P | AL143322 | Tetraodon |
| C 233 | 22 | 1.1 | 601 | 13 | BM114708 | LM0809B04- | C 306 | 1.1 | 864 | 17 | CNS0530J | AL19780 | Tetraodon |
| C 234 | 22 | 1.1 | 601 | 14 | BM729493 | 1h81f12.x | C 307 | 1.1 | 865 | 9 | AL579710 | AL579710 | AL579710 |
| C 235 | 22 | 1.1 | 601 | 14 | BQ015777 | UT-H-DT1- | C 308 | 1.1 | 872 | 17 | AZ528236 | AZ528236 | ENTIDB43TE |
| C 236 | 22 | 1.1 | 603 | 14 | BQ021290 | UT-H-DT1- | C 309 | 1.1 | 883 | 17 | AZ209795 | AZ209795 | SP_0151_B |
| C 237 | 22 | 1.1 | 603 | 14 | BQ086397 | 1h8B08.x | C 310 | 1.1 | 891 | 17 | CNS0305T | AZ525238 | Tetraodon |
| C 238 | 22 | 1.1 | 608 | 17 | AA549982 | 1061m3.9m | C 311 | 1.1 | 893 | 17 | AZ673746 | AZ673746 | ENTIV37TE |
| C 239 | 22 | 1.1 | 608 | 17 | DL16M19S | AL738831 | C 312 | 1.1 | 896 | 17 | CNS015WX | AL165786 | Tetraodon |
| C 240 | 22 | 1.1 | 609 | 17 | AZ305830 | LM0006110 | C 313 | 1.1 | 898 | 13 | BM394843 | BM394843 | 50072-2-6 |
| C 241 | 22 | 1.1 | 610 | 14 | BQ101293 | 1h92f11.x | C 314 | 1.1 | 910 | 17 | AZ204426 | AZ204426 | SP_0057_A |
| C 242 | 22 | 1.1 | 611 | 13 | BM340255 | BM340255 | C 315 | 1.1 | 913 | 13 | BM358329 | BM358329 | GA_Ea000 |
| C 243 | 22 | 1.1 | 611 | 14 | BQ0861376 | 1h90a09.x | C 316 | 1.1 | 941 | 17 | AZ167072 | AZ167072 | SP_0094_B |
| C 244 | 22 | 1.1 | 612 | 10 | AM681376 | EST00094 | C 317 | 1.1 | 942 | 14 | BQ481612 | BQ481612 | PV_GEA005 |
| C 245 | 22 | 1.1 | 613 | 12 | BG085123 | BG085123 | C 318 | 1.1 | 950 | 12 | BG283344 | BG283344 | 602407078 |
| C 246 | 22 | 1.1 | 613 | 14 | BQ086580 | 1h90a09.x | C 319 | 1.1 | 956 | 9 | AL665573 | AL665573 | AL665573 |
| C 247 | 22 | 1.1 | 615 | 14 | BQ272306 | 1j92d12.x | C 320 | 1.1 | 958 | 17 | CNS020CM | AL209119 | Tetraodon |
| C 248 | 22 | 1.1 | 616 | 14 | BQ894378 | AGENCOURT | C 321 | 1.1 | 960 | 17 | AZ204424 | AZ204424 | SP_0057_A |
| C 249 | 22 | 1.1 | 616 | 17 | AZ527184 | AZ527184 | C 322 | 1.1 | 966 | 17 | AZ138425 | AZ138425 | SP_0157_B |
| C 250 | 22 | 1.1 | 617 | 14 | BQ525586 | BQ525586 | C 323 | 1.1 | 967 | 17 | CNS077ZW | AL433832 | T7_end_of |
| C 251 | 22 | 1.1 | 618 | 12 | BG070986 | H3093D10- | C 324 | 1.1 | 972 | 12 | BF133475 | BF133475 | 601646002 |
| C 252 | 22 | 1.1 | 622 | 13 | BM505876 | 1h43B03.x | C 325 | 1.1 | 1003 | 17 | BH156894 | BH156894 | ENTG377R |
| C 253 | 22 | 1.1 | 623 | 14 | BM992598 | UT-H-DT0- | C 326 | 1.1 | 1017 | 12 | BG498181 | BG498181 | 602543569 |
| C 254 | 22 | 1.1 | 625 | 10 | BE258686 | 601116241 | C 327 | 1.1 | 1036 | 17 | CNS033XT | AL295165 | Tetraodon |
| C 255 | 22 | 1.1 | 625 | 14 | BM730800 | 1h61908.x | C 328 | 1.1 | 1038 | 12 | BF790231 | BF790231 | 602249762 |
| C 256 | 22 | 1.1 | 627 | 14 | BQ747337 | UT-M-FA0- | C 329 | 1.1 | 1039 | 17 | CNS0509H | AL155278 | Tetraodon |
| C 257 | 22 | 1.1 | 627 | 14 | AZ724357 | RPC1-24-1 | C 330 | 1.1 | 1101 | 12 | BG572784 | BG572784 | 602594317 |
| C 258 | 22 | 1.1 | 628 | 14 | BQ747050 | UT-M-FA0- | C 331 | 1.1 | 1101 | 17 | CNS00F0U | AL771076 | Drosophila |
| C 259 | 22 | 1.1 | 632 | 14 | BQ006608 | UT-H-EI1- | C 332 | 1.1 | 1119 | 12 | BG677027 | BG677027 | 602623671 |
| C 260 | 22 | 1.1 | 643 | 10 | BB664062 | BB664062 | C 333 | 1.1 | 1238 | 13 | BI490175 | BI490175 | 603031951 |
| C 261 | 22 | 1.1 | 643 | 14 | BQ527424 | NISC.no22 | C 334 | 1.1 | 1242 | 12 | BF120206 | BF120206 | 601756120 |
| C 262 | 22 | 1.1 | 645 | 17 | AZ524462 | AZ524462 | C 335 | 1.1 | 1253 | 12 | BF619400 | BF619400 | HVSMC000 |
| C 263 | 22 | 1.1 | 649 | 14 | BQ563833 | g106909.x | C 336 | 1.1 | 1256 | 12 | BG395337 | BG395337 | 602457872 |
| C 264 | 22 | 1.1 | 652 | 14 | BQ794423 | EST.3361 | C 337 | 1.1 | 1321 | 13 | BM559683 | BM559683 | AGENCOURT |
| C 265 | 22 | 1.1 | 658 | 9 | AL648224 | AL648224 | C 338 | 1.1 | 1444 | 14 | BQ071075 | BQ071075 | AGENCOURT |
| C 266 | 22 | 1.1 | 660 | 10 | AM081121 | XC29d09.x | C 339 | 1.1 | 1465 | 13 | BM476475 | BM476475 | AGENCOURT |
| C 267 | 22 | 1.1 | 667 | 14 | BQ025229 | UT-1-BB1D | C 340 | 1.1 | 1490 | 12 | BG329977 | BG329977 | 602429735 |
| C 268 | 22 | 1.1 | 670 | 17 | AQ751613 | HS.5576.B | C 341 | 1.1 | 1787 | 11 | BC030988 | BC030988 | Homo sapi |
| C 269 | 22 | 1.1 | 671 | 9 | AL317577 | UK61b11.Y | C 342 | 1.1 | 1792 | 12 | BF036146 | BF036146 | 601458153 |
| C 270 | 22 | 1.1 | 678 | 14 | BQ602045 | MI-P-HO-A | C 343 | 1.1 | 74 | 14 | BQ264489 | BQ264489 | NISC_1702 |
| C 271 | 22 | 1.1 | 686 | 14 | BQ016102 | UT-H-DT1- | C 344 | 1.1 | 82 | 12 | BG223123 | BG223123 | nah43909. |
| C 272 | 22 | 1.1 | 687 | 14 | BQ043557 | UT-M-EFO- | C 345 | 1.1 | 91 | 12 | BG027087 | BG027087 | 602293950 |
| C 273 | 22 | 1.1 | 688 | 14 | BQ004657 | UT-H-EIO- | C 346 | 1.1 | 99 | 12 | BG155156 | BG155156 | sab29f02. |
| C 274 | 22 | 1.1 | 690 | 17 | AZ570408 | 274PVB02 | C 347 | 1.1 | 103 | 12 | BF650753 | BF650753 | NF099G09E |
| C 275 | 22 | 1.1 | 691 | 10 | AM012999 | SEEP066K | C 348 | 1.1 | 103 | 13 | BF166321 | BF166321 | CH115B11. |
| C 276 | 22 | 1.1 | 694 | 12 | BM152427 | UT07909.x | C 349 | 1.1 | 120 | 14 | BQ974233 | BQ974233 | CH115B11. |
| C 277 | 22 | 1.1 | 696 | 13 | BM162427 | EST364950 | C 350 | 1.1 | 125 | 13 | BJ434848 | BJ434848 | AL660429 |
| C 278 | 22 | 1.1 | 697 | 14 | BM930669 | UT-E-EJ1- | C 351 | 1.1 | 126 | 9 | AL660429 | AL660429 | OC37121.S |
| C 279 | 22 | 1.1 | 699 | 9 | A0060694 | A0060694 | C 352 | 1.1 | 128 | 9 | AA807306 | AA807306 | BJ076387 |
| C 280 | 22 | 1.1 | 704 | 12 | BF319021 | UY07903.x | C 353 | 1.1 | 132 | 13 | BF1076387 | BF1076387 | 601771026 |
| C 281 | 22 | 1.1 | 705 | 10 | AV721479 | AV721479 | C 354 | 1.1 | 135 | 9 | AL137073 | AL137073 | UT-R-C2B- |
| C 282 | 22 | 1.1 | 706 | 14 | BQ204100 | UT-R-DN1- | C 355 | 1.1 | 136 | 9 | A1561321 | A1561321 | UT-R-C2B- |
| C 283 | 22 | 1.1 | 710 | 14 | BQ018646 | UT-H-DH1- | C 356 | 1.1 | 136 | 9 | A1561321 | A1561321 | UT-R-C2B- |
| C 284 | 22 | 1.1 | 711 | 14 | BQ150149 | UT-R-DH1- | C 357 | 1.1 | 136 | 9 | A1561321 | A1561321 | UT-R-C2B- |
| C 285 | 22 | 1.1 | 716 | 14 | BM971248 | UT-R-DH1- | C 358 | 1.1 | 142 | 9 | A1973949 | A1973949 | sd14b05.Y |
| C 286 | 22 | 1.1 | 720 | 10 | BQ043397 | UT-M-EFO- | C 359 | 1.1 | 147 | 14 | BQ3866289 | BQ3866289 | NISC_lm17 |
| C 287 | 22 | 1.1 | 722 | 10 | AV733303 | AV733303 | C 360 | 1.1 | 149 | 10 | BM056864 | BM056864 | BM056864 |
| C 288 | 22 | 1.1 | 725 | 17 | AG155427 | AG155427 | C 361 | 1.1 | 153 | 9 | AA997003 | AA997003 | UT-R-C0-h |
| C 289 | 22 | 1.1 | 732 | 17 | AG137551 | Pan.troc1 | C 362 | 1.1 | 156 | 10 | BE115960 | BE115960 | UT-R-BST- |
| C 290 | 22 | 1.1 | 734 | 17 | AG187688 | Pan.troc1 | C 363 | 1.1 | 159 | 13 | BM513541 | BM513541 | KX91h02.Y |
| C 291 | 22 | 1.1 | 738 | 10 | BE250181 | 601092949 | C 364 | 1.1 | 159 | 14 | N96086 | N96086 | 2550C3.czap |
| C 292 | 22 | 1.1 | 735 | 10 | AV682099 | AV682099 | C 365 | 1.1 | 164 | 12 | BG3711769 | BG3711769 | UT-R-CV0- |
| C 293 | 22 | 1.1 | 737 | 14 | BM676907 | UT-E-E01- | C 366 | 1.1 | 165 | 9 | AA770165 | AA770165 | ab84d02.s |
| C 294 | 22 | 1.1 | 758 | 17 | AZ137089 | SP_0175.A | C 367 | 1.1 | 165 | 14 | BQ205920 | BQ205920 | UT-R-EPO- |
| C 295 | 22 | 1.1 | 774 | 12 | BG402959 | 602418763 | C 368 | 1.1 | 168 | 9 | A1356764 | A1356764 | GYY22804.x |
| C 296 | 22 | 1.1 | 775 | 14 | BQ174442 | UT-M-DJ2- | C 369 | 1.1 | 170 | 13 | BI083318 | BI083318 | 602875579 |
| C 297 | 22 | 1.1 | 776 | 17 | CNS011HD | AL100267 | C 370 | 1.1 | 174 | 13 | BI294463 | BI294463 | UT-R-DK0- |
| C 298 | 22 | 1.1 | 781 | 10 | AV757286 | AV757286 | C 371 | 1.1 | 186 | 14 | BM528718 | BM528718 | AGENCOURT |

| | | | | | | | | | | | | | | | |
|-------|----|-----|-----|----|----------|----------|-------------|-------|----|-----|-----|----|-----------|-----------|-------------|
| C 372 | 21 | 1.0 | 191 | 10 | BE107773 | BE107773 | UI-R-CA0- | C 445 | 21 | 1.0 | 334 | 9 | A1850149 | A1850149 | UI-M-BG1- |
| C 373 | 21 | 1.0 | 192 | 9 | A1043683 | A1043683 | UI-R-C0-J | C 446 | 21 | 1.0 | 341 | 9 | A1058743 | A1058743 | UI-R-C1-K |
| C 374 | 21 | 1.0 | 194 | 12 | BF794592 | BF794592 | 602255860 | C 447 | 21 | 1.0 | 342 | 10 | BE118034 | BE118034 | UI-R-B51- |
| C 375 | 21 | 1.0 | 195 | 14 | BQ017004 | BQ017004 | UI-H-DT1- | C 448 | 21 | 1.0 | 343 | 10 | AM015716 | AM015716 | UI-R-B10P |
| C 376 | 21 | 1.0 | 200 | 10 | BB482544 | BB482544 | UI-R-B52- | C 449 | 21 | 1.0 | 346 | 12 | BE846405 | BE846405 | EST0068 G |
| C 377 | 21 | 1.0 | 204 | 12 | BF397294 | BF397294 | UI-R-B52- | C 450 | 21 | 1.0 | 347 | 10 | AM069573 | AM069573 | C147C01.x |
| C 378 | 21 | 1.0 | 205 | 9 | A1849244 | A1849244 | UI-M-AJ1- | C 451 | 21 | 1.0 | 347 | 12 | BC377282 | BC377282 | UI-R-CU0- |
| C 379 | 21 | 1.0 | 206 | 13 | BI403941 | BI403941 | MI-P-CF1- | C 452 | 21 | 1.0 | 350 | 10 | AM460238 | AM460238 | sh6h05.y |
| C 380 | 21 | 1.0 | 211 | 9 | A1611185 | A1611185 | UI-M-AJ0- | C 453 | 21 | 1.0 | 350 | 13 | BI280691 | BI280691 | UI-R-CS05 |
| C 381 | 21 | 1.0 | 211 | 9 | A1840314 | A1840314 | UI-M-AJ0- | C 454 | 21 | 1.0 | 351 | 9 | A1850167 | A1850167 | UI-M-BG1- |
| C 382 | 21 | 1.0 | 216 | 9 | A1847145 | A1847145 | UI-M-AJ0- | C 455 | 21 | 1.0 | 351 | 14 | BM899830 | BM899830 | UI-M-DJ1- |
| C 383 | 21 | 1.0 | 216 | 10 | AM122789 | AM122789 | UI-M-BH2- | C 456 | 21 | 1.0 | 353 | 9 | AA856728 | AA856728 | od54C05.s |
| C 384 | 21 | 1.0 | 216 | 13 | BJ087332 | BJ087332 | AGENCOCURT | C 457 | 21 | 1.0 | 353 | 9 | A1045510 | A1045510 | UI-R-C1-K |
| C 385 | 21 | 1.0 | 217 | 14 | BO735764 | BO735764 | AGENCOCURT | C 458 | 21 | 1.0 | 353 | 9 | A1850345 | A1850345 | UI-M-BG1- |
| C 386 | 21 | 1.0 | 219 | 10 | AV534094 | AV534094 | AGENCOCURT | C 459 | 21 | 1.0 | 354 | 9 | A1454876 | A1454876 | UI-R-C2P- |
| C 387 | 21 | 1.0 | 220 | 14 | BO452030 | BO452030 | PFSFOA9 | C 460 | 21 | 1.0 | 356 | 10 | AM628478 | AM628478 | UI-R-C2P- |
| C 388 | 21 | 1.0 | 222 | 12 | BF391570 | BF391570 | UI-R-CA1- | C 461 | 21 | 1.0 | 358 | 9 | A1129050 | A1129050 | ox87f12.s |
| C 389 | 21 | 1.0 | 223 | 9 | A1501012 | A1501012 | UI-R-C2P- | C 462 | 21 | 1.0 | 358 | 10 | BE344944 | BE344944 | h066e10.x |
| C 390 | 21 | 1.0 | 228 | 12 | BG076508 | BG076508 | H3001A05- | C 463 | 21 | 1.0 | 359 | 10 | AV386091 | AV386091 | AV386091 |
| C 391 | 21 | 1.0 | 234 | 10 | BE521912 | BE521912 | M22DX1TM | C 464 | 21 | 1.0 | 362 | 10 | BE542258 | BE542258 | 601066594 |
| C 392 | 21 | 1.0 | 235 | 9 | AA371539 | AA371539 | EST83386 | C 465 | 21 | 1.0 | 363 | 14 | R21519 | R21519 | yh19C02.s1 |
| C 393 | 21 | 1.0 | 240 | 9 | AU074503 | AU074503 | AGENCOCURT | C 466 | 21 | 1.0 | 369 | 9 | AA670379 | AA670379 | ad16q10.s |
| C 394 | 21 | 1.0 | 248 | 12 | BG261275 | BG261275 | 602373020 | C 467 | 21 | 1.0 | 370 | 9 | A1843111 | A1843111 | UI-M-AK1- |
| C 395 | 21 | 1.0 | 249 | 9 | A1491360 | A1491360 | 466017G05 | C 468 | 21 | 1.0 | 375 | 10 | AM060541 | AM060541 | UI-M-BH1- |
| C 396 | 21 | 1.0 | 250 | 9 | A1286312 | A1286312 | 94919411.x | C 469 | 21 | 1.0 | 375 | 14 | N98111 | N98111 | 2201C3 CZAP |
| C 397 | 21 | 1.0 | 252 | 12 | BG261188 | BG261188 | 602372921 | C 470 | 21 | 1.0 | 378 | 10 | AM081003 | AM081003 | xc39B03.x |
| C 398 | 21 | 1.0 | 255 | 10 | AV368747 | AV368747 | UI-R-DK0- | C 471 | 21 | 1.0 | 381 | 10 | AV405631 | AV405631 | AV405631 |
| C 399 | 21 | 1.0 | 256 | 13 | BI294173 | BI294173 | UI-R-DK0- | C 472 | 21 | 1.0 | 382 | 9 | A1844431 | A1844431 | UI-M-AJ1- |
| C 400 | 21 | 1.0 | 259 | 14 | BQ962867 | BQ962867 | AGENCOCURT | C 473 | 21 | 1.0 | 383 | 9 | A1735150 | A1735150 | as68e10.x |
| C 401 | 21 | 1.0 | 264 | 12 | BF702785 | BF702785 | MI-P-H3-a | C 474 | 21 | 1.0 | 383 | 9 | A1782325 | A1782325 | AL782325 |
| C 402 | 21 | 1.0 | 265 | 13 | BI288411 | BI288411 | UI-R-DK0- | C 475 | 21 | 1.0 | 386 | 17 | A2509891 | A2509891 | 1M0354J02 |
| C 403 | 21 | 1.0 | 266 | 14 | BO155158 | BO155158 | NF076H12I | C 476 | 21 | 1.0 | 388 | 10 | AM046068 | AM046068 | UI-M-BH1- |
| C 404 | 21 | 1.0 | 269 | 12 | BF414863 | BF414863 | UI-R-BJ2- | C 477 | 21 | 1.0 | 388 | 14 | N98042 | N98042 | 2104C3 CZAP |
| C 405 | 21 | 1.0 | 272 | 10 | BE177732 | BE177732 | UI-R-BJ2- | C 478 | 21 | 1.0 | 391 | 17 | BH335496 | BH335496 | CH230-5F1 |
| C 406 | 21 | 1.0 | 274 | 12 | BF150137 | BF150137 | uy82604.Y | C 479 | 21 | 1.0 | 392 | 10 | AM199069 | AM199069 | da11b10.x |
| C 407 | 21 | 1.0 | 275 | 17 | A2490693 | A2490693 | 1M0333N18 | C 480 | 21 | 1.0 | 392 | 12 | BC796761 | BC796761 | UI-R-SMSA |
| C 408 | 21 | 1.0 | 276 | 9 | AU010774 | AU010774 | AM010774 | C 481 | 21 | 1.0 | 394 | 9 | A1374818 | A1374818 | ta56h09.x |
| C 409 | 21 | 1.0 | 278 | 12 | BF552106 | BF552106 | UI-R-C2P- | C 482 | 21 | 1.0 | 395 | 17 | A2897218 | A2897218 | RPCT-24-2 |
| C 410 | 21 | 1.0 | 278 | 10 | AM448919 | AM448919 | UI-R-BJ2- | C 483 | 21 | 1.0 | 396 | 9 | A1099749 | A1099749 | 33902 Lam |
| C 411 | 21 | 1.0 | 279 | 10 | BE135646 | BE135646 | UI-R-BJ2- | C 484 | 21 | 1.0 | 401 | 14 | N97805 | N97805 | 1286C3 CZAP |
| C 412 | 21 | 1.0 | 285 | 9 | AJ012994 | AJ012994 | EHEST172 | C 485 | 21 | 1.0 | 403 | 13 | BE1652876 | BE1652876 | 603300011 |
| C 413 | 21 | 1.0 | 285 | 14 | BM966786 | BM966786 | 1j26e06.Y | C 486 | 21 | 1.0 | 404 | 12 | BF703324 | BF703324 | MI-P-A3-a |
| C 414 | 21 | 1.0 | 285 | 14 | BM966902 | BM966902 | 1j27h11.Y | C 487 | 21 | 1.0 | 404 | 14 | BO156369 | BO156369 | NF092B03I |
| C 415 | 21 | 1.0 | 287 | 13 | BI133428 | BI133428 | UI-M-BH3- | C 488 | 21 | 1.0 | 406 | 10 | AV814860 | AV814860 | AV814860 |
| C 416 | 21 | 1.0 | 288 | 12 | BE956094 | BE956094 | UI-M-BH4- | C 489 | 21 | 1.0 | 406 | 12 | BF512209 | BF512209 | UI-R-BI3- |
| C 417 | 21 | 1.0 | 289 | 12 | BG894248 | BG894248 | Kt09c11.Y | C 490 | 21 | 1.0 | 410 | 9 | AA556128 | AA556128 | TENT0467 |
| C 418 | 21 | 1.0 | 291 | 9 | AA371538 | AA371538 | EST83385 | C 491 | 21 | 1.0 | 410 | 10 | AM119563 | AM119563 | sd48a11.Y |
| C 419 | 21 | 1.0 | 294 | 9 | AA134265 | AA134265 | z021909.s | C 492 | 21 | 1.0 | 411 | 9 | AA598749 | AA598749 | ae49h10.s |
| C 420 | 21 | 1.0 | 295 | 10 | BE672732 | BE672732 | 7b75f12.x | C 493 | 21 | 1.0 | 411 | 10 | AV806484 | AV806484 | AV806484 |
| C 421 | 21 | 1.0 | 295 | 14 | N97777 | N97777 | 1244C3 CZAP | C 494 | 21 | 1.0 | 412 | 9 | A1070279 | A1070279 | UI-R-Y0-1 |
| C 422 | 21 | 1.0 | 296 | 12 | BF388505 | BF388505 | UI-R-CA1- | C 495 | 21 | 1.0 | 413 | 10 | AM524076 | AM524076 | UI-R-B00- |
| C 423 | 21 | 1.0 | 296 | 14 | BO359963 | BO359963 | PM4-HN014 | C 496 | 21 | 1.0 | 415 | 9 | A1843611 | A1843611 | UI-M-BH4- |
| C 424 | 21 | 1.0 | 297 | 13 | BI193460 | BI193460 | UI-M-BH3- | C 497 | 21 | 1.0 | 415 | 10 | AM790786 | AM790786 | DO0190-R |
| C 425 | 21 | 1.0 | 297 | 13 | BI133873 | BI133873 | UI-M-BH3- | C 498 | 21 | 1.0 | 415 | 12 | BE956157 | BE956157 | UI-M-BH4- |
| C 426 | 21 | 1.0 | 297 | 13 | BI302724 | BI302724 | UI-R-DK0- | C 499 | 21 | 1.0 | 416 | 9 | AA900776 | AA900776 | UI-R-ED-d |
| C 427 | 21 | 1.0 | 299 | 17 | AQ388177 | AQ388177 | RPC111-14 | C 500 | 21 | 1.0 | 416 | 14 | BO595956 | BO595956 | PFSFOA2 |
| C 428 | 21 | 1.0 | 303 | 13 | BI497443 | BI497443 | df136f07. | C 501 | 21 | 1.0 | 420 | 10 | BE040545 | BE040545 | OE08F07 O |
| C 429 | 21 | 1.0 | 306 | 13 | BM529903 | BM529903 | fw94C07.x | C 502 | 21 | 1.0 | 421 | 9 | A1840479 | A1840479 | UI-M-AJ0- |
| C 430 | 21 | 1.0 | 307 | 9 | A1849406 | A1849406 | UI-M-AJ1- | C 503 | 21 | 1.0 | 421 | 13 | BM666035 | BM666035 | UI-E-DX1- |
| C 431 | 21 | 1.0 | 315 | 17 | AQ091552 | AQ091552 | HS-2208-A | C 504 | 21 | 1.0 | 421 | 13 | BM668969 | BM668969 | UI-E-CX1- |
| C 432 | 21 | 1.0 | 317 | 9 | A1715788 | A1715788 | UI-R-Y0-a | C 505 | 21 | 1.0 | 424 | 10 | AA964294 | AA964294 | EST074997 |
| C 433 | 21 | 1.0 | 317 | 12 | BE807258 | BE807258 | ss14b08.Y | C 506 | 21 | 1.0 | 425 | 17 | BEH020109 | BEH020109 | 16532a.d |
| C 434 | 21 | 1.0 | 317 | 12 | BE990104 | BE990104 | UI-M-BJ1- | C 507 | 21 | 1.0 | 426 | 9 | A1850431 | A1850431 | UI-M-BG1- |
| C 435 | 21 | 1.0 | 318 | 9 | AL366086 | AL366086 | MEBA04E07 | C 508 | 21 | 1.0 | 427 | 10 | AM147205 | AM147205 | da15d12.x |
| C 436 | 21 | 1.0 | 320 | 13 | BI277403 | BI277403 | UI-R-CY0- | C 509 | 21 | 1.0 | 428 | 9 | AA900294 | AA900294 | UI-R-ED-d |
| C 437 | 21 | 1.0 | 323 | 10 | AM137254 | AM137254 | UI-H-BJ1- | C 510 | 21 | 1.0 | 428 | 9 | AL398679 | AL398679 | AL398679 |
| C 438 | 21 | 1.0 | 324 | 10 | BB222932 | BB222932 | BB222932 | C 511 | 21 | 1.0 | 429 | 9 | A1089997 | A1089997 | qb13901.x |
| C 439 | 21 | 1.0 | 325 | 9 | AA552261 | AA552261 | nJ89e11.s | C 512 | 21 | 1.0 | 429 | 12 | BE955988 | BE955988 | UI-M-BH4- |
| C 440 | 21 | 1.0 | 328 | 12 | BF455526 | BF455526 | UI-M-BJ1- | C 513 | 21 | 1.0 | 432 | 14 | BM899688 | BM899688 | UI-M-DJ1- |
| C 441 | 21 | 1.0 | 328 | 13 | BI290274 | BI290274 | UI-R-DK0- | C 514 | 21 | 1.0 | 431 | 9 | AA604853 | AA604853 | no92904.s |
| C 442 | 21 | 1.0 | 330 | 9 | A1839479 | A1839479 | UI-M-AJ0- | C 515 | 21 | 1.0 | 433 | 9 | AT854584 | AT854584 | UI-M-BH0- |
| C 443 | 21 | 1.0 | 330 | 14 | BM938266 | BM938266 | UI-M-BJ1- | C 516 | 21 | 1.0 | 437 | 14 | BQ086380 | BQ086380 | 1j21h10.Y |
| C 444 | 21 | 1.0 | 330 | 14 | BM938266 | BM938266 | UI-M-BJ1- | C 517 | 21 | 1.0 | 437 | 14 | BQ086380 | BQ086380 | 1j21h10.Y |

| | | | | | | | | | | | | | | | | | | | |
|-----|---|-----|----|-----|-----|----|----------|----------|------------|---|-----|----|-----|-----|------|----------|----------|----------|----------|
| 518 | C | 531 | 21 | 1.0 | 437 | 14 | BO101065 | BO101065 | 1423610.Y | C | 591 | 21 | 1.0 | 515 | 9 | AL677043 | AL677043 | AL677043 | AL677043 |
| 519 | C | 519 | 21 | 1.0 | 438 | 9 | AA868323 | AA868323 | CJ22911.S | C | 592 | 21 | 1.0 | 517 | 13 | B1690090 | B1690090 | B1690090 | B1690090 |
| 520 | C | 520 | 21 | 1.0 | 439 | 9 | AJ283359 | AJ283359 | 4A3B-AAA- | C | 593 | 21 | 1.0 | 519 | 9 | A1654062 | A1654062 | A1654062 | A1654062 |
| 521 | C | 521 | 21 | 1.0 | 441 | 10 | AW530155 | AW530155 | UI-R-C4-a | C | 594 | 21 | 1.0 | 519 | 13 | BM273422 | BM273422 | BM273422 | BM273422 |
| 522 | C | 522 | 21 | 1.0 | 442 | 9 | AI299298 | AI299298 | qn43to6.x | C | 595 | 21 | 1.0 | 519 | 14 | BO027701 | BO027701 | BO027701 | BO027701 |
| 523 | C | 523 | 21 | 1.0 | 443 | 14 | N34580 | N34580 | yy14h08.sl | C | 596 | 21 | 1.0 | 519 | 14 | BO209652 | BO209652 | BO209652 | BO209652 |
| 524 | C | 524 | 21 | 1.0 | 443 | 17 | BH753005 | BH753005 | SALK.0196 | C | 597 | 21 | 1.0 | 521 | 9 | AA874886 | AA874886 | AA874886 | AA874886 |
| 525 | C | 525 | 21 | 1.0 | 444 | 9 | AI200821 | AI200821 | q67601.x | C | 598 | 21 | 1.0 | 522 | 9 | A1673365 | A1673365 | A1673365 | A1673365 |
| 526 | C | 526 | 21 | 1.0 | 444 | 14 | BQ264840 | BQ264840 | NISC-1101 | C | 599 | 21 | 1.0 | 522 | 9 | A1669975 | A1669975 | A1669975 | A1669975 |
| 527 | C | 527 | 21 | 1.0 | 446 | 10 | AW355448 | AW355448 | S47DI.AGS | C | 600 | 21 | 1.0 | 524 | 12 | BM373491 | BM373491 | BM373491 | BM373491 |
| 528 | C | 528 | 21 | 1.0 | 447 | 10 | AW362934 | AW362934 | EST375007 | C | 601 | 21 | 1.0 | 525 | 10 | AM028305 | AM028305 | AM028305 | AM028305 |
| 529 | C | 529 | 21 | 1.0 | 448 | 13 | BI802110 | BI802110 | H070B10.E | C | 602 | 21 | 1.0 | 525 | 12 | BE710723 | BE710723 | BE710723 | BE710723 |
| 530 | C | 530 | 21 | 1.0 | 449 | 9 | AL375526 | AL375526 | MEB12C01 | C | 603 | 21 | 1.0 | 525 | 13 | BM265403 | BM265403 | BM265403 | BM265403 |
| 531 | C | 531 | 21 | 1.0 | 449 | 10 | AV738426 | AV738426 | AV738426 | C | 604 | 21 | 1.0 | 525 | 14 | BQ001203 | BQ001203 | BQ001203 | BQ001203 |
| 532 | C | 532 | 21 | 1.0 | 450 | 9 | AI165123 | AI165123 | qe51H03.X | C | 605 | 21 | 1.0 | 525 | 17 | A292834 | A292834 | A292834 | A292834 |
| 533 | C | 533 | 21 | 1.0 | 451 | 10 | AW922826 | AW922826 | UI-H-B12- | C | 606 | 21 | 1.0 | 527 | 12 | BF395412 | BF395412 | BF395412 | BF395412 |
| 534 | C | 534 | 21 | 1.0 | 451 | 13 | BG913965 | BG913965 | 602810624 | C | 607 | 21 | 1.0 | 529 | 9 | AI134440 | AI134440 | AI134440 | AI134440 |
| 535 | C | 535 | 21 | 1.0 | 451 | 14 | BQ677130 | BQ677130 | AGENCOURT | C | 608 | 21 | 1.0 | 531 | 12 | BF401492 | BF401492 | BF401492 | BF401492 |
| 536 | C | 536 | 21 | 1.0 | 453 | 13 | BI597456 | BI597456 | 603243673 | C | 609 | 21 | 1.0 | 532 | 12 | BE147615 | BE147615 | BE147615 | BE147615 |
| 537 | C | 537 | 21 | 1.0 | 453 | 9 | AL653469 | AL653469 | AL653469 | C | 610 | 21 | 1.0 | 532 | 14 | BO200751 | BO200751 | BO200751 | BO200751 |
| 538 | C | 538 | 21 | 1.0 | 454 | 17 | AA550649 | AA550649 | NP054G08F | C | 611 | 21 | 1.0 | 532 | 17 | A2895259 | A2895259 | A2895259 | A2895259 |
| 539 | C | 539 | 21 | 1.0 | 454 | 13 | BI270767 | BI270767 | q296106.X | C | 612 | 21 | 1.0 | 534 | 10 | BE101834 | BE101834 | BE101834 | BE101834 |
| 540 | C | 540 | 21 | 1.0 | 458 | 9 | AI373565 | AI373565 | UI-M-AJ0- | C | 613 | 21 | 1.0 | 534 | 14 | BQ159424 | BQ159424 | BQ159424 | BQ159424 |
| 541 | C | 541 | 21 | 1.0 | 458 | 9 | AI838385 | AI838385 | 602143775 | C | 614 | 21 | 1.0 | 535 | 10 | BE097984 | BE097984 | BE097984 | BE097984 |
| 542 | C | 542 | 21 | 1.0 | 459 | 12 | BF686395 | BF686395 | UI-H-B12- | C | 615 | 21 | 1.0 | 535 | 10 | BE111827 | BE111827 | BE111827 | BE111827 |
| 543 | C | 543 | 21 | 1.0 | 460 | 10 | AW291333 | AW291333 | UI-M-BH2- | C | 616 | 21 | 1.0 | 540 | 12 | BF583686 | BF583686 | BF583686 | BF583686 |
| 544 | C | 544 | 21 | 1.0 | 461 | 10 | AW079061 | AW079061 | xb46H03.X | C | 617 | 21 | 1.0 | 541 | 17 | AQ340290 | AQ340290 | AQ340290 | AQ340290 |
| 545 | C | 545 | 21 | 1.0 | 462 | 10 | AW222214 | AW222214 | UI-M-BH2- | C | 618 | 21 | 1.0 | 542 | 9 | AL801344 | AL801344 | AL801344 | AL801344 |
| 546 | C | 546 | 21 | 1.0 | 464 | 10 | AW296657 | AW296657 | UI-H-B12- | C | 619 | 21 | 1.0 | 545 | 12 | BF968512 | BF968512 | BF968512 | BF968512 |
| 547 | C | 547 | 21 | 1.0 | 464 | 10 | BE114198 | BE114198 | UI-R-C4-a | C | 620 | 21 | 1.0 | 548 | 10 | AV575423 | AV575423 | AV575423 | AV575423 |
| 548 | C | 548 | 21 | 1.0 | 465 | 9 | AI793799 | AI793799 | f054C05.X | C | 621 | 21 | 1.0 | 549 | 12 | BE954522 | BE954522 | BE954522 | BE954522 |
| 549 | C | 549 | 21 | 1.0 | 465 | 10 | AV699149 | AV699149 | AV699149 | C | 622 | 21 | 1.0 | 548 | 12 | BE952805 | BE952805 | BE952805 | BE952805 |
| 550 | C | 550 | 21 | 1.0 | 465 | 12 | BE511087 | BE511087 | UI-H-B12- | C | 623 | 21 | 1.0 | 550 | 10 | AW561709 | AW561709 | AW561709 | AW561709 |
| 551 | C | 551 | 21 | 1.0 | 469 | 12 | BE956817 | BE956817 | UI-M-BG2- | C | 624 | 21 | 1.0 | 551 | 12 | BR395517 | BR395517 | BR395517 | BR395517 |
| 552 | C | 552 | 21 | 1.0 | 470 | 13 | BM319819 | BM319819 | k178608.Y | C | 625 | 21 | 1.0 | 553 | 9 | A1678445 | A1678445 | A1678445 | A1678445 |
| 553 | C | 553 | 21 | 1.0 | 470 | 17 | A2574980 | A2574980 | 334PVH08 | C | 626 | 21 | 1.0 | 557 | 13 | B1509208 | B1509208 | B1509208 | B1509208 |
| 554 | C | 554 | 21 | 1.0 | 471 | 9 | AI841773 | AI841773 | UI-M-ANO- | C | 627 | 21 | 1.0 | 557 | 17 | A2281266 | A2281266 | A2281266 | A2281266 |
| 555 | C | 555 | 21 | 1.0 | 472 | 9 | AI027143 | AI027143 | 0w41912.S | C | 628 | 21 | 1.0 | 561 | 13 | BI740656 | BI740656 | BI740656 | BI740656 |
| 556 | C | 556 | 21 | 1.0 | 472 | 10 | AW451425 | AW451425 | UI-H-B13- | C | 629 | 21 | 1.0 | 563 | 13 | BM539679 | BM539679 | BM539679 | BM539679 |
| 557 | C | 557 | 21 | 1.0 | 472 | 14 | N26478 | N26478 | yx61D06.sl | C | 630 | 21 | 1.0 | 564 | 14 | BO601009 | BO601009 | BO601009 | BO601009 |
| 558 | C | 558 | 21 | 1.0 | 478 | 12 | BE975500 | BE975500 | bs43F02.X | C | 631 | 21 | 1.0 | 565 | 9 | AI229684 | AI229684 | AI229684 | AI229684 |
| 559 | C | 559 | 21 | 1.0 | 485 | 10 | AM014087 | AM014087 | UI-H-B10- | C | 632 | 21 | 1.0 | 568 | 13 | BM386810 | BM386810 | BM386810 | BM386810 |
| 560 | C | 560 | 21 | 1.0 | 485 | 14 | BQ193689 | BQ193689 | UI-R-CN1- | C | 633 | 21 | 1.0 | 568 | 14 | BQ195706 | BQ195706 | BQ195706 | BQ195706 |
| 561 | C | 561 | 21 | 1.0 | 486 | 14 | BQ523808 | BQ523808 | NISC-nc01 | C | 634 | 21 | 1.0 | 569 | 14 | BM677119 | BM677119 | BM677119 | BM677119 |
| 562 | C | 562 | 21 | 1.0 | 487 | 13 | BI224160 | BI224160 | 602942431 | C | 635 | 21 | 1.0 | 569 | 14 | BO017444 | BO017444 | BO017444 | BO017444 |
| 563 | C | 563 | 21 | 1.0 | 488 | 14 | BQ538469 | BQ538469 | MEST601-F | C | 636 | 21 | 1.0 | 570 | 9 | AU261996 | AU261996 | AU261996 | AU261996 |
| 564 | C | 564 | 21 | 1.0 | 488 | 17 | A2447131 | A2447131 | 1M0243823 | C | 637 | 21 | 1.0 | 570 | 14 | BO600782 | BO600782 | BO600782 | BO600782 |
| 565 | C | 565 | 21 | 1.0 | 489 | 14 | BO520351 | BO520351 | NISC.n104 | C | 638 | 21 | 1.0 | 571 | 14 | BM725593 | BM725593 | BM725593 | BM725593 |
| 566 | C | 566 | 21 | 1.0 | 491 | 9 | AA900741 | AA900741 | UI-R-E0-D | C | 639 | 21 | 1.0 | 572 | 14 | BM971844 | BM971844 | BM971844 | BM971844 |
| 567 | C | 567 | 21 | 1.0 | 491 | 12 | BE956352 | BE956352 | UI-M-BH4- | C | 640 | 21 | 1.0 | 574 | 9 | AA390731 | AA390731 | AA390731 | AA390731 |
| 568 | C | 568 | 21 | 1.0 | 491 | 14 | BQ394783 | BQ394783 | NISC.nc911 | C | 641 | 21 | 1.0 | 574 | 10 | BE209833 | BE209833 | BE209833 | BE209833 |
| 569 | C | 569 | 21 | 1.0 | 492 | 12 | BE556733 | BE556733 | UI-R-C0-h | C | 642 | 21 | 1.0 | 574 | 14 | BI841680 | BI841680 | BI841680 | BI841680 |
| 570 | C | 570 | 21 | 1.0 | 493 | 10 | AM003318 | AM003318 | wq65D03.x | C | 643 | 21 | 1.0 | 574 | 17 | AO962789 | AO962789 | AO962789 | AO962789 |
| 571 | C | 571 | 21 | 1.0 | 497 | 14 | BO800086 | BO800086 | EST.7121 | C | 644 | 21 | 1.0 | 575 | 9 | AL627797 | AL627797 | AL627797 | AL627797 |
| 572 | C | 572 | 21 | 1.0 | 499 | 9 | AI336174 | AI336174 | ms33G05.X | C | 645 | 21 | 1.0 | 576 | 9 | A1737326 | A1737326 | A1737326 | A1737326 |
| 573 | C | 573 | 21 | 1.0 | 499 | 9 | AI845832 | AI845832 | UI-M-AOI | C | 646 | 21 | 1.0 | 576 | 12 | BE956867 | BE956867 | BE956867 | BE956867 |
| 574 | C | 574 | 21 | 1.0 | 499 | 10 | AM717413 | AM717413 | f141C09.X | C | 647 | 21 | 1.0 | 577 | 14 | BM678068 | BM678068 | BM678068 | BM678068 |
| 575 | C | 575 | 21 | 1.0 | 499 | 12 | BE991240 | BE991240 | UI-M-B21- | C | 648 | 21 | 1.0 | 581 | 13 | BM483648 | BM483648 | BM483648 | BM483648 |
| 576 | C | 576 | 21 | 1.0 | 500 | 9 | AU086134 | AU086134 | AU086134 | C | 649 | 21 | 1.0 | 585 | 10 | BB587923 | BB587923 | BB587923 | BB587923 |
| 577 | C | 577 | 21 | 1.0 | 500 | 9 | AU086142 | AU086142 | AU086142 | C | 650 | 21 | 1.0 | 585 | 14 | BO272379 | BO272379 | BO272379 | BO272379 |
| 578 | C | 578 | 21 | 1.0 | 500 | 9 | AU086896 | AU086896 | AU086896 | C | 651 | 21 | 1.0 | 587 | 9 | A1336019 | A1336019 | A1336019 | A1336019 |
| 579 | C | 579 | 21 | 1.0 | 500 | 9 | AU086985 | AU086985 | AU086985 | C | 652 | 21 | 1.0 | 587 | 10 | AV699700 | AV699700 | AV699700 | AV699700 |
| 580 | C | 580 | 21 | 1.0 | 500 | 9 | AU087237 | AU087237 | AU087237 | C | 653 | 21 | 1.0 | 588 | 12 | BE956851 | BE956851 | BE956851 | BE956851 |
| 581 | C | 581 | 21 | 1.0 | 500 | 9 | AU087541 | AU087541 | AU087541 | C | 654 | 21 | 1.0 | 589 | 13 | BQ205122 | BQ205122 | BQ205122 | BQ205122 |
| 582 | C | 582 | 21 | 1.0 | 500 | 9 | AU088379 | AU088379 | AU088379 | C | 655 | 21 | 1.0 | 592 | 14 | BO600546 | BO600546 | BO600546 | BO600546 |
| 583 | C | 583 | 21 | 1.0 | 502 | 12 | BF564120 | BF564120 | UI-R-C4-a | C | 656 | 21 | 1.0 | 593 | 13 | BI167345 | BI167345 | BI167345 | BI167345 |
| 584 | C | 584 | 21 | 1.0 | 503 | 10 | AM123187 | AM123187 | UI-M-BH2- | C | 657 | 21 | 1.0 | 596 | 10 | AM116293 | AM116293 | AM116293 | AM116293 |
| 585 | C | 585 | 21 | 1.0 | 505 | 9 | AI716226 | AI716226 | UI-R-Y0-a | C | 658 | 21 | 1.0 | 599 | 10</ | | | | |

| | | | | | | | | | | | | | | | |
|-----|----|-----|-----|----|-----------|-----------|------------|-------|----|-----|------|----|-----------|-----------|------------|
| 664 | 21 | 1.0 | 611 | 14 | BQ071830 | BQ071830 | AGENCOURT | C 737 | 21 | 1.0 | 732 | 14 | BQ207100 | BQ207100 | UT-R-D21- |
| 665 | 21 | 1.0 | 612 | 13 | BM479588 | BM479588 | AGENCOURT | C 738 | 21 | 1.0 | 735 | 9 | AA820710 | AA820710 | LD24567.5 |
| 666 | 21 | 1.0 | 613 | 13 | BI171784 | BI171784 | 602933843 | C 739 | 21 | 1.0 | 735 | 14 | BQ183544 | BQ183544 | UT-H-EUO- |
| 667 | 21 | 1.0 | 613 | 14 | BQ952457 | BQ952457 | AGENCOURT | C 740 | 21 | 1.0 | 735 | 14 | BQ576028 | BQ576028 | UT-H-E21- |
| 668 | 21 | 1.0 | 614 | 10 | AM188650 | AM188650 | XK01F11.X | C 741 | 21 | 1.0 | 737 | 14 | BQ725873 | BQ725873 | UT-H-FHO- |
| 669 | 21 | 1.0 | 615 | 14 | BQ395314 | BQ395314 | NISC_mg14 | C 742 | 21 | 1.0 | 737 | 17 | AL189339 | AL189339 | SP-1014.A |
| 670 | 21 | 1.0 | 617 | 14 | BQ709366 | BQ709366 | AGENCOURT | C 743 | 21 | 1.0 | 742 | 13 | BI837648 | BI837648 | 603086869 |
| 671 | 21 | 1.0 | 622 | 12 | BG842014 | BG842014 | MEST35-A0 | C 744 | 21 | 1.0 | 743 | 14 | BQ193266 | BQ193266 | UT-R-DRI- |
| 672 | 21 | 1.0 | 623 | 13 | BI777443 | BI777443 | UT-R-CY0- | C 745 | 21 | 1.0 | 746 | 14 | BQ201359 | BQ201359 | UT-R-D01- |
| 673 | 21 | 1.0 | 623 | 17 | AO969547 | AO969547 | LERJ016TF | C 746 | 21 | 1.0 | 755 | 14 | BQ005866 | BQ005866 | UT-H-ED1- |
| 674 | 21 | 1.0 | 624 | 14 | BQ210772 | BQ210772 | UT-R-DY1- | C 747 | 21 | 1.0 | 755 | 17 | BH266279 | BH266279 | CH230-19A |
| 675 | 21 | 1.0 | 626 | 13 | BI17531 | BI17531 | 603041818 | C 748 | 21 | 1.0 | 757 | 10 | BE256021 | BE256021 | 601113536 |
| 676 | 21 | 1.0 | 628 | 9 | AA843950 | AA843950 | AK05501.S | C 749 | 21 | 1.0 | 762 | 17 | AL2878438 | AL2878438 | R0CT-23-1 |
| 677 | 21 | 1.0 | 628 | 14 | BQ598826 | BQ598826 | MI-P-E4-a | C 750 | 21 | 1.0 | 764 | 14 | BM977651 | BM977651 | UT-CF-EN1 |
| 678 | 21 | 1.0 | 630 | 17 | BH669275 | BH669275 | BOHTE53TR | C 751 | 21 | 1.0 | 766 | 14 | BM977727 | BM977727 | UT-CF-EN1 |
| 679 | 21 | 1.0 | 630 | 12 | BG286349 | BG286349 | 602383120 | C 752 | 21 | 1.0 | 766 | 17 | AG073616 | AG073616 | Par. trogl |
| 680 | 21 | 1.0 | 631 | 17 | BH079110 | BH079110 | RPCI-24-2 | C 753 | 21 | 1.0 | 776 | 17 | CNS05GKE | CNS05GKE | |
| 681 | 21 | 1.0 | 632 | 11 | BC013839 | BC013839 | Home sapi | C 754 | 21 | 1.0 | 779 | 12 | BE965697 | BE965697 | 601679638 |
| 682 | 21 | 1.0 | 632 | 11 | BM510522 | BM510522 | 1J46G08.X | C 755 | 21 | 1.0 | 785 | 10 | AM013477 | AM013477 | PC075KS.W |
| 683 | 21 | 1.0 | 633 | 14 | BQ221996- | BQ221996- | AGENCOURT | C 756 | 21 | 1.0 | 785 | 17 | CNS0457T | CNS0457T | telradodon |
| 684 | 21 | 1.0 | 633 | 14 | BQ398561 | BQ398561 | NISC_m008 | C 757 | 21 | 1.0 | 792 | 12 | BF108207 | BF108207 | 601824281 |
| 685 | 21 | 1.0 | 633 | 17 | BO711589 | BO711589 | AGENCOURT | C 758 | 21 | 1.0 | 797 | 12 | BG500712 | BG500712 | 602547117 |
| 686 | 21 | 1.0 | 633 | 17 | AZ995163 | AZ995163 | 2M0280619 | C 759 | 21 | 1.0 | 798 | 17 | AG111112 | AG111112 | Par. trogl |
| 687 | 21 | 1.0 | 634 | 13 | BM944721 | BM944721 | 50072-2-5 | C 760 | 21 | 1.0 | 812 | 17 | CNS00IDC | CNS00IDC | |
| 688 | 21 | 1.0 | 636 | 9 | AL629037 | AL629037 | AL629037 | C 761 | 21 | 1.0 | 822 | 17 | CNS06DK1 | CNS06DK1 | |
| 689 | 21 | 1.0 | 640 | 12 | BE51736 | BE51736 | UT-M-CCO- | C 762 | 21 | 1.0 | 838 | 17 | AL189945 | AL189945 | SP-0179-LB |
| 690 | 21 | 1.0 | 641 | 12 | BM514896 | BM514896 | AGENCOURT | C 763 | 21 | 1.0 | 847 | 17 | AQ075199 | AQ075199 | GA-5569-LB |
| 691 | 21 | 1.0 | 641 | 14 | BF508649 | BF508649 | UI-H-B14- | C 764 | 21 | 1.0 | 857 | 17 | CNS02MWL | CNS02MWL | |
| 692 | 21 | 1.0 | 641 | 17 | BH768511 | BH768511 | BMBA360D | C 765 | 21 | 1.0 | 858 | 12 | BF107950 | BF107950 | 601823978 |
| 693 | 21 | 1.0 | 642 | 10 | BE113122 | BE113122 | UI-R-BJ1- | C 766 | 21 | 1.0 | 871 | 12 | BG564376 | BG564376 | 602584303 |
| 694 | 21 | 1.0 | 642 | 10 | BE577061 | BE577061 | 7d91e09.X | C 767 | 21 | 1.0 | 871 | 12 | BE903478 | BE903478 | 601676727 |
| 695 | 21 | 1.0 | 643 | 10 | AV695123 | AV695123 | AV695123 | C 768 | 21 | 1.0 | 880 | 12 | BF273314 | BF273314 | GA-EB001 |
| 696 | 21 | 1.0 | 643 | 17 | AZ556210 | AZ556210 | 2M0222508 | C 769 | 21 | 1.0 | 885 | 12 | BE887076 | BE887076 | 601506564 |
| 697 | 21 | 1.0 | 644 | 14 | BQ525821 | BQ525821 | NISC-ng12 | C 770 | 21 | 1.0 | 902 | 9 | BQ947021 | BQ947021 | |
| 698 | 21 | 1.0 | 646 | 10 | AV682770 | AV682770 | AV682770 | C 771 | 21 | 1.0 | 907 | 14 | AL547152 | AL547152 | AL547152 |
| 699 | 21 | 1.0 | 646 | 14 | BQ907046 | BQ907046 | MO24C09.O | C 772 | 21 | 1.0 | 934 | 14 | BQ954928 | BQ954928 | |
| 700 | 21 | 1.0 | 646 | 17 | AO968572 | AO968572 | LERJ042TR | C 773 | 21 | 1.0 | 935 | 12 | BF274499 | BF274499 | GA-EB002 |
| 701 | 21 | 1.0 | 647 | 13 | BI734265 | BI734265 | UI-R-CW0- | C 774 | 21 | 1.0 | 954 | 12 | EG075101 | EG075101 | H3143C10- |
| 702 | 21 | 1.0 | 650 | 14 | BQ447234 | BQ447234 | UI-H-ED1- | C 775 | 21 | 1.0 | 979 | 17 | AL157764 | AL157764 | SP-0058.A |
| 703 | 21 | 1.0 | 652 | 14 | BQ227244 | BQ227244 | AGENCOURT | C 776 | 21 | 1.0 | 981 | 17 | CNS06DUD | CNS06DUD | |
| 704 | 21 | 1.0 | 655 | 14 | BQ796228 | BQ796228 | EST-5166 | C 777 | 21 | 1.0 | 982 | 13 | BM416142 | BM416142 | |
| 705 | 21 | 1.0 | 656 | 14 | BQ781798 | BQ781798 | UI-R-EF0- | C 778 | 21 | 1.0 | 982 | 17 | CNS05ZEO | CNS05ZEO | |
| 706 | 21 | 1.0 | 657 | 13 | BI503737 | BI503737 | BI170022B | C 779 | 21 | 1.0 | 996 | 9 | AL564765 | AL564765 | AL564765 |
| 707 | 21 | 1.0 | 657 | 14 | BM984860 | BM984860 | UI-CF-EC1 | C 780 | 21 | 1.0 | 1012 | 12 | CNS06CMM | CNS06CMM | |
| 708 | 21 | 1.0 | 658 | 14 | BQ202480 | BQ202480 | UI-R-EB1- | C 781 | 21 | 1.0 | 1028 | 12 | BG65167 | BG65167 | 602784134 |
| 709 | 21 | 1.0 | 659 | 13 | BI296712 | BI296712 | UI-R-CV2- | C 782 | 21 | 1.0 | 1031 | 12 | BF526526 | BF526526 | 602070834 |
| 710 | 21 | 1.0 | 661 | 10 | AV822633 | AV822633 | AV822633 | C 783 | 21 | 1.0 | 1034 | 17 | CNS02OVS | CNS02OVS | |
| 711 | 21 | 1.0 | 661 | 14 | BQ043696 | BQ043696 | UI-M-EF0- | C 784 | 21 | 1.0 | 1043 | 17 | CNS05N70 | CNS05N70 | |
| 712 | 21 | 1.0 | 662 | 12 | BG776589 | BG776589 | 602663731 | C 785 | 21 | 1.0 | 1066 | 12 | BF036758 | BF036758 | |
| 713 | 21 | 1.0 | 662 | 14 | BQ747244 | BQ747244 | UI-M-FA0- | C 786 | 21 | 1.0 | 1082 | 17 | AG086541 | AG086541 | Par. trogl |
| 714 | 21 | 1.0 | 663 | 17 | AZ566381 | AZ566381 | 220PvH01 | C 787 | 21 | 1.0 | 1114 | 13 | BI462948 | BI462948 | 603204153 |
| 715 | 21 | 1.0 | 664 | 13 | BI503842 | BI503842 | BI170027B | C 788 | 21 | 1.0 | 1117 | 14 | BQ424219 | BQ424219 | |
| 716 | 21 | 1.0 | 670 | 12 | BG380713 | BG380713 | UI-R-CT0- | C 789 | 21 | 1.0 | 1169 | 14 | BM809466 | BM809466 | |
| 717 | 21 | 1.0 | 673 | 14 | BQ399973 | BQ399973 | NISC_m006 | C 790 | 21 | 1.0 | 1248 | 17 | AG059113 | AG059113 | |
| 718 | 21 | 1.0 | 675 | 10 | BB665076 | BB665076 | BB665076 | C 791 | 21 | 1.0 | 1274 | 14 | BQ654234 | BQ654234 | |
| 719 | 21 | 1.0 | 676 | 14 | BQ018481 | BQ018481 | UI-H-DH1- | C 792 | 21 | 1.0 | 1356 | 12 | BG326380 | BG326380 | |
| 720 | 21 | 1.0 | 676 | 14 | BQ574444 | BQ574444 | UI-H-E21- | C 793 | 21 | 1.0 | 2217 | 11 | BC022916 | BC022916 | Mus. muscu |
| 721 | 21 | 1.0 | 679 | 14 | BQ185797 | BQ185797 | UI-E-EJ1- | C 794 | 21 | 1.0 | 76 | 12 | BF049433 | BF049433 | |
| 722 | 21 | 1.0 | 681 | 14 | BM990364 | BM990364 | UI-H-DH0- | C 795 | 21 | 1.0 | 81 | 9 | AT254169 | AT254169 | qV54e05.X |
| 723 | 21 | 1.0 | 686 | 14 | BM009395 | BM009395 | UI-H-ED1- | C 796 | 21 | 1.0 | 81 | 9 | AT254603 | AT254603 | qV54e12.X |
| 724 | 21 | 1.0 | 688 | 14 | BQ005566 | BQ005566 | UI-H-ET1- | C 797 | 21 | 1.0 | 82 | 9 | AT254283 | AT254283 | qV33912.X |
| 725 | 21 | 1.0 | 690 | 17 | AG162799 | AG162799 | Par. trogl | C 798 | 21 | 1.0 | 86 | 9 | AT287932 | AT287932 | qU26102.X |
| 726 | 21 | 1.0 | 691 | 17 | AA550226 | AA550226 | 1346m3.gm | C 799 | 21 | 1.0 | 87 | 14 | BQ400167 | BQ400167 | NISC_m007 |
| 727 | 21 | 1.0 | 692 | 9 | AU176190 | AU176190 | AU176190 | C 800 | 21 | 1.0 | 92 | 12 | BG058287 | BG058287 | na123b10. |
| 728 | 21 | 1.0 | 693 | 13 | BI274534 | BI274534 | UI-R-CW0- | C 801 | 21 | 1.0 | 98 | 12 | BG151274 | BG151274 | na135910. |
| 729 | 21 | 1.0 | 694 | 9 | AU176229 | AU176229 | AU176229 | C 802 | 21 | 1.0 | 100 | 12 | BF726152 | BF726152 | by01407.X |
| 730 | 21 | 1.0 | 696 | 13 | BM390834 | BM390834 | UI-R-CN1- | C 803 | 21 | 1.0 | 100 | 12 | BG058212 | BG058212 | na123b10. |
| 731 | 21 | 1.0 | 697 | 14 | BM978618 | BM978618 | UI-CF-D01 | C 804 | 21 | 1.0 | 101 | 12 | BG272015 | BG272015 | na161f06. |
| 732 | 21 | 1.0 | 699 | 14 | BM981867 | BM981867 | UI-CF-EN1 | C 805 | 21 | 1.0 | 113 | 12 | BG231424 | BG231424 | na142e12. |
| 733 | 21 | 1.0 | 700 | 14 | BQ199676 | BQ199676 | UI-R-D01- | C 806 | 21 | 1.0 | 117 | 13 | BM090148 | BM090148 | 505291.MA |
| 734 | 21 | 1.0 | 708 | 9 | AA439842 | AA439842 | LD14518.5 | C 807 | 21 | 1.0 | 117 | 13 | BM484619 | BM484619 | 538539.MA |
| 735 | 21 | 1.0 | 719 | 14 | BM993245 | BM993245 | UI-H-DT0- | C 808 | 21 | 1.0 | 117 | 14 | BQ096868 | BQ096868 | MO20608.O |
| 736 | 21 | 1.0 | 728 | 10 | BE309725 | BE309725 | 601095817 | C 809 | 21 | 1.0 | 120 | 10 | AW354931 | AW354931 | 37588.MAR |

| | | | | | | | | | | | | | |
|-------|----|-----|-----|----|-----------|----------------------|-------|----|-----|-----|----|-----------|----------------------|
| C 810 | 20 | 1.0 | 122 | 9 | A1872554 | A1872554 wmt75907.x | C 883 | 20 | 1.0 | 254 | 9 | A1189443 | A1189443 qd02809.x |
| C 811 | 20 | 1.0 | 122 | 10 | BE682760 | BE682760 181079.MA | C 884 | 20 | 1.0 | 254 | 12 | BF653319 | BF653319 277122.MA |
| C 812 | 20 | 1.0 | 131 | 10 | BF387908 | BF387908 UT-R-CA1- | C 885 | 20 | 1.0 | 254 | 13 | BM150024 | BM150024 TCAPAD11 |
| C 813 | 20 | 1.0 | 132 | 14 | BQ593635 | BQ593635 PESTOab1 | C 886 | 20 | 1.0 | 255 | 10 | AA374048 | AA374048 EST86320 |
| C 814 | 20 | 1.0 | 136 | 9 | AT036206 | AT036206 wh20c12.x | C 887 | 20 | 1.0 | 255 | 10 | AV783156 | AV783156 AV783156 |
| C 815 | 20 | 1.0 | 139 | 9 | AT311969 | AT311969 ta89e01.x | C 888 | 20 | 1.0 | 260 | 13 | BM095063 | BM095063 sa125h07. |
| C 816 | 20 | 1.0 | 141 | 9 | A1250015 | A1250015 qx24905.x | C 889 | 20 | 1.0 | 261 | 10 | AA380688 | AA380688 EST3927. |
| C 817 | 20 | 1.0 | 144 | 14 | BQ265053 | BQ265053 NISC_L103 | C 890 | 20 | 1.0 | 261 | 10 | AM534716 | AM534716 UT-R-B50- |
| C 818 | 20 | 1.0 | 146 | 13 | BM534103 | BM534103 tw93909.y | C 891 | 20 | 1.0 | 262 | 9 | A1583642 | A1583642 t174004.x |
| C 819 | 20 | 1.0 | 148 | 10 | BE071413 | BE071413 RC2-HT050 | C 892 | 20 | 1.0 | 262 | 9 | AA299688 | AA299688 EST12233 |
| C 820 | 20 | 1.0 | 148 | 12 | BF002054 | BF002054 7498d05.x | C 893 | 20 | 1.0 | 262 | 9 | AA433767 | AA433767 nh02b12.s |
| C 821 | 20 | 1.0 | 151 | 17 | AZ257451 | AZ257451 RPTC1-23-I | C 894 | 20 | 1.0 | 262 | 13 | BM144357 | BM144357 TCAPAD110 |
| C 822 | 20 | 1.0 | 157 | 9 | AA198802 | AA198802 mml1d10.r | C 895 | 20 | 1.0 | 265 | 10 | BE317804 | BE317804 NF070G12L |
| C 823 | 20 | 1.0 | 158 | 13 | BF1344339 | BF1344339 372786.MA | C 896 | 20 | 1.0 | 268 | 13 | BM362814 | BM362814 BS3200500 |
| C 824 | 20 | 1.0 | 159 | 12 | BF418502 | BF418502 UT-R-BJ2- | C 897 | 20 | 1.0 | 268 | 13 | BM364996 | BM364996 BS3200080 |
| C 825 | 20 | 1.0 | 160 | 9 | A1345714 | A1345714 tB84C09.x | C 898 | 20 | 1.0 | 271 | 14 | BM774160 | BM774160 ra90b07.y |
| C 826 | 20 | 1.0 | 161 | 9 | A1252073 | A1252073 qv58a10.x | C 899 | 20 | 1.0 | 272 | 14 | L38218 | L38218 BNAF1100E.M |
| C 827 | 20 | 1.0 | 162 | 12 | BG654215 | BG654215 sad53c05. | C 900 | 20 | 1.0 | 273 | 10 | AM487664 | AM487664 89994.MAR |
| C 828 | 20 | 1.0 | 164 | 9 | A1836277 | A1836277 UT-M-A00- | C 901 | 20 | 1.0 | 277 | 12 | BF581312 | BF581312 602100587 |
| C 829 | 20 | 1.0 | 164 | 14 | BQ739491 | BQ739491 PESTOab4 | C 902 | 20 | 1.0 | 278 | 9 | AA122813 | AA122813 mt03810.r |
| C 830 | 20 | 1.0 | 165 | 9 | A1342993 | A1342993 qoc27e08.x | C 903 | 20 | 1.0 | 278 | 10 | BM445348 | BM445348 BB445348 |
| C 831 | 20 | 1.0 | 166 | 13 | BM157733 | BM157733 tw26f05.y | C 904 | 20 | 1.0 | 279 | 10 | AM524153 | AM524153 UT-R-B00- |
| C 832 | 20 | 1.0 | 168 | 10 | AM292058 | AM292058 UT-H-B12- | C 905 | 20 | 1.0 | 280 | 13 | BM089829 | BM089829 503682.MA |
| C 833 | 20 | 1.0 | 169 | 10 | AV523525 | AV523525 AV523525 | C 906 | 20 | 1.0 | 281 | 14 | BQ769330 | BQ769330 PESTOab1 |
| C 834 | 20 | 1.0 | 174 | 10 | AM273389 | AM273389 x138f09.x | C 907 | 20 | 1.0 | 282 | 12 | BE991525 | BE991525 UT-R-B21- |
| C 835 | 20 | 1.0 | 174 | 10 | AM551273 | AM551273 L0076B08- | C 908 | 20 | 1.0 | 282 | 12 | BF992125 | BF992125 UT-M-B21- |
| C 836 | 20 | 1.0 | 176 | 10 | AM275671 | AM275671 x039f05.x | C 909 | 20 | 1.0 | 283 | 14 | C91311 | C91311 C91311 Dict |
| C 837 | 20 | 1.0 | 177 | 9 | A1434121 | A1434121 t131d06.x | C 910 | 20 | 1.0 | 286 | 10 | AM103667 | AM103667 xe82c07.x |
| C 838 | 20 | 1.0 | 177 | 9 | AL697773 | AL697773 DKFZP6860 | C 911 | 20 | 1.0 | 286 | 12 | BF704545 | BF704545 MI-P-H4-a |
| C 839 | 20 | 1.0 | 179 | 12 | BG182541 | BG182541 RST1417.A | C 912 | 20 | 1.0 | 288 | 9 | A1680307 | A1680307 tw62c02.x |
| C 840 | 20 | 1.0 | 182 | 9 | AU057285 | AU057285 AU057285 | C 913 | 20 | 1.0 | 289 | 12 | BC382642 | BC382642 298614.MA |
| C 841 | 20 | 1.0 | 183 | 9 | AA808666 | AA808666 ob38e09.s | C 914 | 20 | 1.0 | 289 | 14 | BM929517 | BM929517 ut30c11.x |
| C 842 | 20 | 1.0 | 186 | 9 | AA433304 | AA433304 TEMP0127 | C 915 | 20 | 1.0 | 290 | 9 | A1793937 | A1793937 w130c11.x |
| C 843 | 20 | 1.0 | 186 | 10 | BE120123 | BE120123 UT-R-CA0- | C 916 | 20 | 1.0 | 290 | 12 | BF412860 | BF412860 UT-R-B11- |
| C 844 | 20 | 1.0 | 187 | 13 | BJ080118 | BJ080118 BJ080118 | C 917 | 20 | 1.0 | 290 | 13 | BM540016 | BM540016 hb17a08.9 |
| C 845 | 20 | 1.0 | 188 | 12 | BF781021 | BF781021 60210650 | C 918 | 20 | 1.0 | 290 | 14 | BQ184202 | BQ184202 UT-E-EJ1- |
| C 846 | 20 | 1.0 | 190 | 12 | BG627020 | BG627020 CC-est1f1c. | C 919 | 20 | 1.0 | 293 | 10 | AM150288 | AM150288 x937c08.x |
| C 847 | 20 | 1.0 | 193 | 12 | BG361584 | BG361584 qb55a08.y | C 920 | 20 | 1.0 | 294 | 12 | BC3272147 | BC3272147 na163d10. |
| C 848 | 20 | 1.0 | 193 | 13 | BM144239 | BM144239 TCAPAD110 | C 921 | 20 | 1.0 | 295 | 9 | AA210268 | AA210268 mu7b10.0.r |
| C 849 | 20 | 1.0 | 197 | 12 | BF512879 | BF512879 UT-H-B13- | C 922 | 20 | 1.0 | 295 | 9 | A1915281 | A1915281 x46e03.x |
| C 850 | 20 | 1.0 | 200 | 14 | BQ596238 | BQ596238 PESTOab3 | C 923 | 20 | 1.0 | 296 | 14 | N96010 | N96010 2033c3 c2AP |
| C 851 | 20 | 1.0 | 202 | 13 | BI494079 | BI494079 d1107e09. | C 924 | 20 | 1.0 | 297 | 14 | BQ194687 | BQ194687 UT-R-CM1- |
| C 852 | 20 | 1.0 | 204 | 13 | BJ518521 | BJ518521 BJ518521 | C 925 | 20 | 1.0 | 298 | 17 | BM426986 | BM426986 BOHME29TF |
| C 853 | 20 | 1.0 | 207 | 10 | AV366356 | AV366356 AV366356 | C 926 | 20 | 1.0 | 300 | 9 | AL1373462 | AL1373462 MEB01A01 |
| C 854 | 20 | 1.0 | 207 | 10 | AM122482 | AM122482 UT-M-BH2- | C 927 | 20 | 1.0 | 300 | 9 | AU057349 | AU057349 AU057349 |
| C 855 | 20 | 1.0 | 212 | 9 | A1454761 | A1454761 UT-R-C2P- | C 928 | 20 | 1.0 | 301 | 9 | AA996895 | AA996895 UT-R-C0-h |
| C 856 | 20 | 1.0 | 212 | 13 | BG941967 | BG941967 ax19e07.x | C 929 | 20 | 1.0 | 301 | 14 | BM125882 | BM125882 ko46g05.y |
| C 857 | 20 | 1.0 | 217 | 13 | BI508797 | BI508797 BI170008A | C 930 | 20 | 1.0 | 303 | 10 | BE108291 | BE108291 UT-R-B51- |
| C 858 | 20 | 1.0 | 217 | 13 | BM193679 | BM193679 TCAPAD113 | C 931 | 20 | 1.0 | 303 | 12 | BC626902 | BC626902 CC-est1f1cL |
| C 859 | 20 | 1.0 | 217 | 14 | BQ577165 | BQ577165 PESTOab1 | C 932 | 20 | 1.0 | 304 | 9 | A1282038 | A1282038 q988e10.x |
| C 860 | 20 | 1.0 | 218 | 12 | BG571976 | BG571976 60252430 | C 933 | 20 | 1.0 | 304 | 13 | BI508033 | BI508033 BI170006B |
| C 861 | 20 | 1.0 | 218 | 14 | H21316 | H21316 AS31SB231SL | C 934 | 20 | 1.0 | 306 | 12 | BF403103 | BF403103 UT-R-CA0- |
| C 862 | 20 | 1.0 | 221 | 12 | BE990056 | BE990056 UT-M-B21- | C 935 | 20 | 1.0 | 307 | 9 | AU057452 | AU057452 AU057452 |
| C 863 | 20 | 1.0 | 221 | 12 | BE990056 | BE990056 UT-M-B21- | C 936 | 20 | 1.0 | 307 | 17 | CNS01NV3 | CNS01NV3 ANOpheles |
| C 864 | 20 | 1.0 | 224 | 9 | BB589593 | BB589593 BB589593 | C 937 | 20 | 1.0 | 308 | 10 | BE158660 | BE158660 RCL-HT037 |
| C 865 | 20 | 1.0 | 228 | 9 | A1342140 | A1342140 q126d06.x | C 938 | 20 | 1.0 | 308 | 14 | BQ754571 | BQ754571 EBedc01.S0 |
| C 866 | 20 | 1.0 | 228 | 14 | BM812127 | BM812127 K144901.Y | C 939 | 20 | 1.0 | 308 | 14 | BQ770646 | BQ770646 UT-M-F10- |
| C 867 | 20 | 1.0 | 231 | 12 | BF389709 | BF389709 UT-R-B52- | C 940 | 20 | 1.0 | 310 | 9 | AL373463 | AL373463 MEB01A01 |
| C 868 | 20 | 1.0 | 233 | 13 | BI849908 | BI849908 477882.MA | C 941 | 20 | 1.0 | 311 | 13 | BI153086 | BI153086 GM22156.5 |
| C 869 | 20 | 1.0 | 235 | 10 | AM397585 | AM397585 sg81f04.y | C 942 | 20 | 1.0 | 311 | 14 | BO152676 | BO152676 NF032C05I |
| C 870 | 20 | 1.0 | 241 | 10 | BE156861 | BE156861 RCL-HT037 | C 943 | 20 | 1.0 | 312 | 14 | BO152676 | BO152676 NF019CC05I |
| C 871 | 20 | 1.0 | 242 | 9 | AV263129 | AV263129 AV263129 | C 944 | 20 | 1.0 | 313 | 14 | BQ393435 | BQ393435 NISC-ng03 |
| C 872 | 20 | 1.0 | 243 | 10 | AM452160 | AM452160 UT-H-B13- | C 945 | 20 | 1.0 | 316 | 10 | AV567190 | AV567190 AV567190 |
| C 873 | 20 | 1.0 | 243 | 10 | AM493776 | AM493776 UT-M-BH3- | C 946 | 20 | 1.0 | 316 | 10 | AM291148 | AM291148 UT-H-B12- |
| C 874 | 20 | 1.0 | 244 | 10 | BE057197 | BE057197 sm98a11.y | C 947 | 20 | 1.0 | 317 | 9 | BI823031 | BI823031 603039631 |
| C 875 | 20 | 1.0 | 246 | 9 | A1031183 | A1031183 5MTP.Xeno | C 948 | 20 | 1.0 | 317 | 9 | AI448797 | AI448797 ms39h08.x |
| C 876 | 20 | 1.0 | 248 | 10 | BE201696 | BE201696 FK97d11.x | C 949 | 20 | 1.0 | 320 | 9 | AA779839 | AA779839 a445e09.s |
| C 877 | 20 | 1.0 | 249 | 9 | A1483140 | A1483140 v136c06.x | C 950 | 20 | 1.0 | 320 | 10 | AV738188 | AV738188 AV738188 |
| C 878 | 20 | 1.0 | 249 | 12 | BE995637 | BE995637 UT-M-C0GP | C 951 | 20 | 1.0 | 321 | 10 | AM168414 | AM168414 x182h01.x |
| C 879 | 20 | 1.0 | 249 | 17 | N97395 | N97395 0562M7.gmbp | C 952 | 20 | 1.0 | 321 | 12 | BG290837 | BG290837 602389179 |
| C 880 | 20 | 1.0 | 251 | 10 | AM520963 | AM520963 UT-R-B00- | C 953 | 20 | 1.0 | 322 | 10 | AM122680 | AM122680 UT-M-BH2- |
| C 881 | 20 | 1.0 | 251 | 14 | BQ926880 | BQ926880 AGENCOURT | C 954 | 20 | 1.0 | 323 | 9 | A1333674 | A1333674 q908h04.x |
| C 882 | 20 | 1.0 | 253 | 9 | A1577328 | A1577328 UT-R-Y0-V | C 955 | 20 | 1.0 | 324 | 9 | AA643802 | AA643802 np10e06.s |

RESULT 1
 BI823850
 LOCUS
 DEFINITION
 BI823850 954 bp MRNA linear EST 04-OCT-2000
 603038012F1 NIH_MGC_115 Homo sapiens CDNA clone IMAGE:5179826 5',
 mRNA sequence.
 BI823850
 BI823850.1 GI:15935400
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 954)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@email.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Invitae Genomics, Inc.
 Clone distribution: MGC clone distribution information can be

| Query Match | Best Local Similarity | Matches 630; Conservative | 31.0%; Score 630; DB 13; Length 954; 100.0%; Pred. No. 1.le-186; |
|-------------|---|---------------------------|--|
| 160 | CAAGGCTCTCCACTTCAGACAAATGCTACTAGCTCTAAATAAAGAGAGCTGCTGCTTTCT | 56 | CAAGGCTCTCCACTTCAGACAAATGCTACTAGCTCTAAATAAAGAGAGCTGCTGCTTTCT |
| 116 | GGTTTCCCGAGAGGCGCATCTGGCTCTATCTTCCACATGTACAGCCACAACATCAGCTC | 220 | GGTTTCCCGAGAGGCGCATCTGGCTCTATCTTCCACATGTACAGCCACAACATCAGCTC |
| 280 | CCGTGTATGAAGGACACAGCCCGAGAGCGATCAGATGCTGCTTCTGCTTCTCTGGGCGCTC | 176 | CCGTGTATGAAGGACACAGCCCGAGAGCGATCAGATGCTGCTTCTGCTTCTCTGGGCGCTC |
| 340 | TGGCTCTCTTTTGGGGGACAGCTTTTGGGACAGACCCAGATGTTTCTTACCTGATGGA | 236 | TGGCTCTCTTTTGGGGGACAGCTTTTGGGACAGACCCAGATGTTTCTTACCTGATGGA |
| 400 | GCGCGCTGGACAGTGTGGATGACCTTCAAGACAGAGCACCGCGCTGATGCTGACATGGC | | |

RESULT 2
BM969292/c 593 bp mRNA linear EST 21-MAR-2002
LOCUS
DEFINITION UI-CF-ENO-acp-1-21-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
ACCESSION BM969292
VERSION BM969292.1 GI:19586879
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 593)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704447
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel.: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLY-A=Yes.
FEATURES
SOURCE
Location/Qualifiers
1..593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-ENO-acp-1-21-0-UI"
/clone_1db="UI-CF-ENO"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (11 phage resistant)"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: Ecor I; Site_2: Not I;
UI-CF-ENO is a cDNA library containing the following
tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells.
The library was constructed according to Bonaldo, Lennon
and Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an Ecor I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (d)T18 tail. The sequence tag for this
library is CTGCTCAGGT.
TAG-LIB-UI-CF-ENO
TAG-TISSUE-Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_SEQ-CTGCTCAGGT"
BASE COUNT 157 a 127 c 137 g 172 t
ORIGIN
Query Match 28.6%; Score 581; DB 14; Length 593;
Best Local Similarity 100.0%; Pred. No. 2.7e-171;
Matches 581; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 1436 ACATCTGTGGGTATCAGACTGAGTGTGTCACACGTCGTCAGCAGAGGA 1495
|||||
DB 593 ACATCTGTGGGTATCAGACTGAGTGTGTCACACGTCGTCAGCAGAGGA 534
QY 1436 CTTTGTGTCATGCTTGTGTAGAAAAACAGACTGGGAACTTATGTGAGCAGACAT 1555

|||||
DB 533 CTTTGTGTCATGCTTGTGTAGAAAAACAGACTGGGAACTTATGTGAGCAGACAT 474
QY 1556 CCCACAGTGAACAGAGGATATGCTCTTCTTTCTTGATCTTCCTGTCGGCAGAC 1615
|||||
DB 473 CCCACAGTGAACAGAGGATATGCTCTTCTTTCTTGATCTTCCTGTCGGCAGAC 414
QY 1616 TTGAGAGACTTTGGGCTGAGAGCCATTTAAGCAGACACATATCAGTGGATGATC 1675
|||||
DB 413 TTGAGAGACTTTGGGCTGAGAGCCATTTAAGCAGACACATATCAGTGGATGATC 354
QY 1676 CATTAACCTCCCTGTCACATCTTGCCCAATGGGATGATTTCCACCAAGAGCTCA 1735
|||||
DB 353 CATTAACCTCCCTGTCACATCTTGCCCAATGGGATGATTTCCACCAAGAGCTCA 294
QY 1736 CCAGCATTTTCCACAGAGATGACAAATTTGAGCCCTTGGAGTTCCAGTGGATTCAGGA 1795
|||||
DB 293 CCAGCATTTTCCACAGAGATGACAAATTTGAGCCCTTGGAGTTCCAGTGGATTCAGGA 234
QY 1796 AGGAGTGGGAACAGAGTTGATGCCCTACTTATGAGCTTACCATCAGACTATCGGTAA 1855
|||||
DB 233 AGGAGTGGGAACAGAGTTGATGCCCTACTTATGAGCTTACCATCAGACTATCGGTAA 174
QY 1856 TCAGAAATATGAACAAATCTCTGCACAAAGAGCAGCTTTAAGTTCCAGGGTGGC 1915
|||||
DB 173 TCAGAAATATGAACAAATCTCTGCACAAAGAGCAGCTTTAAGTTCCAGGGTGGC 114
QY 1916 TGGGCTCATTTGATATCTCTCCCTGCTGCAATTTTCCATACATAGAGACTTTGAC 1975
|||||
DB 113 TGGGCTCATTTGATATCTCTCCCTGCTGCAATTTTCCATACATAGAGACTTTGAC 54
QY 1976 CTGGAAGCTGCATCTGTTAATCTAAATTCCTCAATTA 2016
|||||
DB 53 CTGGAAGCTGCATCTGTTAATCTAAATTCCTCAATTA 13
RESULT 3
BM129080 571 bp mRNA linear EST 12-MAR-2002
LOCUS
DEFINITION h1f7c04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
CDNA clone IMAGE:5676655 5', similar to PR:Q9Y5R3 Q9Y5R3
N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
ACCESSION BM129080
VERSION BM129080.1 GI:17123632
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 571)
Melton,D., Brown,J., Kenty,G., Permut,A., Lee,C., Kaestner,K.,
Lemishka,I., Seearce,M., Brestelli,J., Gradowh,J., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blustein,A.,
Schmitt,A., Theising,B., Ritzer,E., Renko,I., Bennett,J., Cardenas
M., Gibbons,M., McCann,R., Cole,R., Tsagaratshvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other_ESTS: h1f7c04.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 434.

FEATURES
SOURCELocation/Qualifiers
1. 571

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:567655"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"

/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 109 a 169 c 143 g 150 t
ORIGIN

Query Match 27 1%; Score 551; DB 13; Length 571;
Best Local Similarity 100.0%; Pred. No. 6,1e-162;
Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 TGAAGTCAGAGAGTAAAGCTTACCTTACAGCTTCTGGAGCGAGTGTCTTC 147
DB 1 TGAGTCTAGAGAGTAAAGCTTACCTTACAGCTTCTGGAGCGAGTGTCTTC 60
QY 148 AAGCCGCTTGCAGAGCTTCCACTTCAGACATCTACTGCTTAAAGTGAAGCT 207
DB 61 AAGCCGCTTGCAGAGCTTCCACTTCAGACATCTACTGCTTAAAGTGAAGCT 120
QY 208 CCGTGTCTTGTGTTCCAGATGAGCATCTTGGCTCTATCTTCCACATGTACAGCA 267
DB 121 CCGTGTCTTGTGTTCCAGATGAGCATCTTGGCTCTATCTTCCACATGTACAGCA 180
QY 268 CAACATCAGCTCCCTGTCTATGAAGCAGACGCCGAGCGATGACGTGCTTCTGTC 327
DB 181 CAACATCAGCTCCCTGTCTATGAAGCAGACGCCGAGCGATGACGTGCTTCTGTC 240
QY 328 TTCCGAGCGCTCGCTCTTCTTGTGGGAGCCTTTTGGGAGCAGCCAGATGTTT 387
DB 241 TTCCGAGCGCTCGCTCTTCTTGTGGGAGCCTTTTGGGAGCAGCCAGATGTTT 300
QY 388 CTACCTGATGAGCGCCGCTGAGCAGTGTGATGACCTTAAGCAGACGCCGCTGAT 447
DB 301 CTACCTGATGAGCGCCGCTGAGCAGTGTGATGACCTTAAGCAGACGCCGCTGAT 360
QY 448 GCTCAGATGAGCTGTGGGAGATGATAGCGCGCTTCTTGTGCGACATGAGCGTCT 507
DB 361 GCTCAGATGAGCTGTGGGAGATGATAGCGCGCTTCTTGTGCGACATGAGCGTCT 420
QY 508 TGATGCTTACATGAGACCTGATCCCGGAGACAGTCCAGCCTTTTCATGGGAGAACAG 567
DB 421 TGATGCTTACATGAGACCTGATCCCGGAGACAGTCCAGCCTTTTCATGGGAGAACAG 480
QY 568 CCGGCGCTGTGTTCTGACACTGCTGTGACATCATCCACAAGATCAATCATCCCCG 627
DB 481 CCGGCGCTGTGTTCTGACACTGCTGTGACATCATCCACAAGATCAATCATCCCCG 540
QY 628 GGCTCAGTCA 638
DB 541 GGCTCAGTCA 551

RESULT 4

BM128370/c
LOCUS BM128370 553 bp mRNA linear EST 12-MAR-2002
DEFINITION i13d12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
ACCESSION CDNA clone IMAGE:5676335 3', mRNA sequence.
VERSION BM128370
KEYWORDS BM128370.1 GI:17122922
SOURCE EST.
ORGANISM human.
Human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 553)
AUTHORS Melton,D., Brown,J., Kently,G., Pettmutt,A., Lee,C., Keestner,K.,
Lemishka,I., Scaerice,M., Brestcailly,J., Gradwohl,G., Clifton,S.,
Hillier,L., Maria,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Reisling,B., Rafter,E., Ronko,I., Bennett,J., Cardenas
M., Gibbons,M., McCann,R., Cole,R., Tsagarisvill,R., Williams,T.,
Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brownjfas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES
SOURCELocation/Qualifiers
1. 553

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5676335"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGIN

Query Match 23 5%; Score 478; DB 13; Length 553;
Best Local Similarity 100.0%; Pred. No. 3,3e-139;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1540 TATGAGAGACACATCCACAGTGAACAGGATATGCTCTTCTTCTTCTGATCT 1599
DB 553 TATGAGAGACACATCCACAGTGAACAGGATATGCTCTTCTTCTTCTGATCT 494
QY 1600 TCCGTCTGGGAGACTTCAGAGACTTGTGCGCTGAGAGCCATTAAAGCAGACACT 1659
DB 493 TCCGTCTGGGAGACTTCAGAGACTTGTGCGCTGAGAGCCATTAAAGCAGACACT 434

QY 1660 ATCACTGGAATTCATCCATAAACCCTCCCTGTCCACATCTTGGCCCAATGGGAATGATCT 1719
 |||||||
 Db 433 ATCACTGGAATTCATCCATAAACCCTCCCTGTCCACATCTTGGCCCAATGGGAATGATCT 374
 |||||||
 QY 1720 TTCACCAAGAGCTACACAGCATTTTCCACAGAGATGCAATTTCTGAGCCCTTGAGTTC 1779
 |||||||
 Db 373 TTCACCAAGAGCTACACAGCATTTTCCACAGAGATGCAATTTCTGAGCCCTTGAGTTC 314
 |||||||
 QY 1780 CCAGTGGATTCAAGGAGGAAGTGGGAACAAGTTGATGCTTACTTATGAGCTTGACCA 1839
 |||||||
 Db 313 CCAGTGGATTCAAGGAGGAAGTGGGAACAAGTTGATGCTTACTTATGAGCTTGACCA 254
 |||||||
 QY 1840 TCACAGCATGCTATGAGAAATATGAACAATCTCTGCACAAAAGACCAAGCTCTT 1899
 |||||||
 Db 253 TCACAGCATGCTATGAGAAATATGAACAATCTCTGCACAAAAGACCAAGCTCTT 194
 |||||||
 QY 1900 AAGTTCACAGGAGTGGCTGGCTGCATTTGATATACATCTCCCTCTGCATTTTCCATCA 1959
 |||||||
 Db 193 AAGTTCACAGGAGTGGCTGGCTGCATTTGATATACATCTCCCTCTGCATTTTCCATCA 134
 |||||||
 QY 1960 CATAGAAACCTTTGACCTGTGAGCTGCCATCTGTATTAATAATTCCCAATTAAG 2017
 |||||||
 Db 133 CATAGAAACCTTTGACCTGTGAGCTGCCATCTGTATTAATAATTCCCAATTAAG 76
 |||||||

RESULT 5

BM128831/c

LOCUS 553 bp mRNA linear EST 12-MAR-2002

DEFINITION 117C04.X1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens

CDNA clone IMAGE:5676655 3', mRNA sequence.

ACCESSION BM128831

VERSION BM128831.1 GI:17123383

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 553)

MELTON,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,

Lemisha,I., Scares,M., Brestelli,J., Gradwohl,G., Clifton,S.,

Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,T., Blisstein,A.,

Schmitt,A., Theising,B., Kitter,E., Ronko,I., Bennett,T., Cardenas

,M., Gibbons,M., McCann,R., Cole,R., Tsagaris,I., Williams,T.,

Jackson,Y. and Powers,Y.

Endocrine Pancreas Consortium

Unpublished (2000)

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@ionp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:

Washington University Genome Sequencing Center for information on

obtaining a clone please contact: Juliana Brown

(brown@fas.harvard.edu) This sequence now available from the IMAGE

consortium, for clone orders contact: info@image.llnl.gov

High quality sequence stop: 425.

FEATURES

SOURCE

1..553
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5676655"
 /clone_id="Melton Normalized Human Islet 4 N4-HIS 1"
 /sex="Both"
 /tissue-type="Islets of Langerhans"
 /dev_stage="Adult"
 /lab_host="DH10B"
 /note="Organ: Pancreas; Vector: pSPORT1; Site:1: Not 1;
 Site_2: Sal 1; Starting library constructed using
 SuperScript Plasmid Library kit (Life Technologies). cDNA

made by oligo-dt priming. Size-selected by column
 fractionation; average insert size 1.08 kb. Library was
 amplified once on solid support and plasmid DNA from
 library was prepared. The library DNA was normalized by
 method #4 from Bonaldo, Lennon, and Soares 1996 Genome
 Research 6:791-806; 0.5 microgram single-stranded library
 plasmid DNA was mixed with 5 micrograms PCR product
 representing library inserts and hybridized to an Ecot of
 20. Single-stranded (unhybridized) plasmids were isolated
 by hydroxyapatite chromatography and used to make this
 library."

BASE COUNT

ORIGIN

Query Match 23.5%; Score 478; DB 13; Length 553;

Best Local Similarity 100.0%; Pred. No. 3.3e-139;

Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1540 TATGTGACGACGACATCCACGACGATGTAACAGGATATGCTCTTCTTTCTGATCT 1599
 |||||||
 Db 553 TATGTGACGACGACATCCACGACGATGTAACAGGATATGCTCTTCTTTCTGATCT 494
 |||||||
 QY 1600 TCCTGTGTGGGACAGACTTCAGAGACTTTGTGGCTGGAGGACTTATAGCAGCAGAGT 1659
 |||||||
 Db 493 TCCTGTGTGGGACAGACTTCAGAGACTTTGTGGCTGGAGGACTTATAGCAGCAGAGT 434
 |||||||
 QY 1660 ATCACTGGAATTCATCCATAAACCCTCCCTGTCCACATCTTGGCCCAATGGGAATGATCT 1719
 |||||||
 Db 433 ATCACTGGAATTCATCCATAAACCCTCCCTGTCCACATCTTGGCCCAATGGGAATGATCT 374
 |||||||
 QY 1720 TTCACCAAGAGCTACACAGCATTTTCCACAGAGATGCAATTTCTGAGCCCTTGAGTTC 1779
 |||||||
 Db 373 TTCACCAAGAGCTACACAGCATTTTCCACAGAGATGCAATTTCTGAGCCCTTGAGTTC 314
 |||||||
 QY 1780 CCAGTGGATTCAAGGAGGAAGTGGGAACAAGTTGATGCTTACTTATGAGCTTGACCA 1839
 |||||||
 Db 313 CCAGTGGATTCAAGGAGGAAGTGGGAACAAGTTGATGCTTACTTATGAGCTTGACCA 254
 |||||||
 QY 1840 TCACAGCATGCTATGAGAAATATGAACAATCTCTGCACAAAAGACCAAGCTCTT 1899
 |||||||
 Db 253 TCACAGCATGCTATGAGAAATATGAACAATCTCTGCACAAAAGACCAAGCTCTT 194
 |||||||
 QY 1900 AAGTTCACAGGAGTGGCTGGCTGCATTTGATATACATCTCCCTCTGCATTTTCCATCA 1959
 |||||||
 Db 193 AAGTTCACAGGAGTGGCTGGCTGCATTTGATATACATCTCCCTCTGCATTTTCCATCA 134
 |||||||
 QY 1960 CATAGAAACCTTTGACCTGTGAGCTGCCATCTGTATTAATAATTCCCAATTAAG 2017
 |||||||
 Db 133 CATAGAAACCTTTGACCTGTGAGCTGCCATCTGTATTAATAATTCCCAATTAAG 76
 |||||||

RESULT 6

AL709927

LOCUS 668 bp mRNA linear EST 22-MAR-2002

DEFINITION DKEF268602364.F1 686 (synonym: h1cc3) Homo sapiens cDNA clone

DKEF268602364 5', mRNA sequence.

ACCESSION AL709927

VERSION AL709927.1 GI:19693282

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 668)

Duesterhoeft,A., Lauder,J., Mewes,H.W., Gassenhuber,J. and Wiemann

EST (Duesterhoeft, et al.)

Unpublished (1999)

Contact: Duesterhoeft A

MIPS Am Klopferstr. 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp68662364) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Source

ggg. . T

/organism="Homo sapiens"

/db_xref="taxon:9606"

```
/clone="DKFZp68602364"
```

/clone lib="586 (synonym: h1cc3)"

/tissue type="human

`udev stage="adult"`

```

/acc-post= "DHI0BT
)ab host=

```

/ IAD_HOUC- DHI0B

ORIGIN

| | | | | |
|-----------------------|-----------------|---------------------|-----------|-------------|
| Query Match | 23.0%; | Score 468; | DB 9; | Length 668; |
| Best Local Similarity | 100.0%; | Pred. No. 3.7e-136; | | |
| Matches 468; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0 |

| | | | |
|----|-----|--|-----|
| QY | 160 | CAAGGCTTCCACATTCGACACATCTCTACTCTCTAAAAAAATGAAGCTCGTGGTTTCT | 219 |
| Db | 72 | CAAGGCTTCCACATTCGACACATCTCTACTCTCTAAAAAAATGAAGCTCGTGGTTTCT | 131 |
| QY | 220 | GGTTTCCACGATGGCATCTTGECTCTATTTCTTCCATGTACAGCCACAATCAGCTC | 279 |
| Db | 132 | GGTTTCCACGATGGCATCTTGECTCTATTTCTTCCATGTACAGCCACAATCAGCTC | 191 |
| QY | 280 | CGTGTATGAAGGACACAGCCGAGCAGTACAGTGGTGTGTCTTCCTCGGGCCTC | 339 |
| Db | 192 | CGTGTATGAAGGACACAGCCGAGCAGTACAGTGGTGTGTCTTCCTCGGGCCTC | 251 |
| QY | 340 | TGSCCTTTCTTTTGTGGGSCACCTTTTGGGSCAGCACCACAATCTTTTCTACTGATGA | 399 |
| Db | 252 | TGSCCTTTCTTTTGTGGGSCACCTTTTGGGSCAGCACCACAATCTTTTCTACTGATGA | 311 |
| QY | 400 | GCCCCCTTGGCACGTGTGATGACCTTCACAGACAGCACCGCCTGATGCTGCACATGAC | 459 |
| Db | 312 | GCCCCCTTGGCACGTGTGATGACCTTCACACAGACAGCACCGCCTGATGCTGCACATGAC | 371 |
| QY | 460 | TGTGGGGAATCTGATACGGGGCGCTTCTTGTGGGACATGAGGCTTTATGCTTACAT | 519 |
| Db | 372 | TGTGGGGAATCTGATACGGGGCGCTTCTTGTGGGACATGAGGCTTTATGCTTACAT | 431 |
| QY | 520 | GGAACCTGTGCCCGAGACAGATCCAGCCTCTTTCAGTGGGAGAACAGCCGGGCGCTGTG | 579 |
| Db | 432 | GGAACCTGTGCCCGAGACAGATCCAGCCTCTTTCAGTGGGAGAACAGCCGGGCGCTGTG | 491 |
| QY | 580 | TTCGTGACCTGCGCTGTGACATCATCTCCACAAGATGAATCATCTCCCCG | 627 |
| Db | 492 | TTCGTGACCTGCGCTGTGACATCATCTCCACAAGATGAATCATCTCCCCG | 539 |

RESULT 7
BF878439 680 bp mRNA EST: 17-JAN-2001
DEFINITION MRO-ET0109-191100-002-h06 ET0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF878439
VERSION BF878439.1 GI:12268569
KEYWORDS EST.
SOURCE human.

TITLE
JOURNAL
MEDLINE
COMMENT

Brustein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, I.F., de Souza, S.J., and Simpson, A.J.C.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.C.

| FEATURES | | | | |
|-----------------------|--|--------------------|------------|-------------|
| source | Location/Qualifiers | | | |
| | 1. 680 | | | |
| | /organism="Homo sapiens" | | | |
| | /db_xref="taxon:9606" | | | |
| | /clone_lib="ET0109" | | | |
| | /dev_stage="Adult" | | | |
| | /note="Organ: lung; tumor: Vector: puc18; site: 1: Smar; site:2: Smar; A mini-library was made by cloning products derived from ORESPRS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." | | | |
| BASE COUNT | 168 a | 164 c | 163 g | 184 t |
| ORIGIN | | | | 1 others |
| Query Match | 15.68; | Score 318; | DB 12; | Length 680; |
| Best Local Similarity | 100.0%; | Pred. No. 1,8e+09; | | |
| Matches 318; | Conservative | 0; | Mismatches | 0; |
| | | | Indels | 0; |
| | | | Gaps | 0; |

[illegible]

| | | | | | |
|------------|-------------------------|--------------|------|--------------------|-----------------|
| RESULT 8 | | | | | |
| AW002418/c | | | | | |
| LOCUS | AW002418 | 362 bp | mRNA | linear | EST 27-OCY-1999 |
| DEFINITION | wu61903.x1 NCI_CGAP_GC6 | Homo sapiens | CDNA | clone IMAGE:552564 | 3', |
| | mRNA sequence. | | | | |
| ACCESSION | AW002418 | | | | |

VERSION AM002418.1 GI:5849334
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 362)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNI at:
www-bio.11nl.gov/ldrf/image/image.html
Seq primer: -400p from G1bco.
Location/Qualifiers
1..362
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NCI-CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI-CGAP_GC4 was prepared, and
ss circles were made in vitro. Following Hsp purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 108 a 69 c 77 g 108 t
ORIGIN
Query Match 14.2%; Score 288; DB 10; Length 362;
Best Local Similarity 100.0%; Pred. No. 6.3e-80;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1730 AGCTACACAGATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGTCCAGTGATT 1789
DB 362 AGCTACACAGATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGTCCAGTGATT 303
QY 1790 CAAGGAGAGAGTGGGAGCAAGGTGGATGCTTACTTATGAGCTTGACATCAGAGCTAT 1849
DB 302 CAAGGAGAGAGTGGGAGCAAGGTGGATGCTTACTTATGAGCTTGACATCAGAGCTAT 243
QY 1850 CGGTATATAGAAATATGAAACAAATCTCTGACACAAAAGAGCAAGCTTTAAGTTCACAG 1909
DB 242 CGGTATATAGAAATATGAAACAAATCTCTGACACAAAAGAGCAAGCTTTAAGTTCACAG 183
QY 1910 GGTGCGTGGGCTGATTTGAATATCCTCCCTCTGCAATTTCCATCAGATGAGAAGAC 1969
DB 182 GGTGCGTGGGCTGATTTGAATATCCTCCCTCTGCAATTTCCATCAGATGAGAAGAC 123
QY 1970 TTGACCTGTGAAGTGCATCTGTTAATTAATAATTCCTCAATTAAG 2017
DB 122 TTGACCTGTGAAGTGCATCTGTTAATTAATAATTCCTCAATTAAG 75
RESULT 9
AM572390/c 419 bp mRNA linear EST 13-MAR-2000
LOCUS x091a06.x2 NCI-CGAP_Col4 Homo sapiens cDNA clone IMAGE:2799634 3'
DEFINITION mRNA sequence.

ACCESSION AM572390
VERSION AM572390.1 GI:7237123
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 419)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNI at:
image.11nl.gov/image/html/resources.shtml
Seq primer: -400p from G1bco
High quality sequence stop: 343.
Location/Qualifiers
1..419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:2799634"
/clone_lib="NCI-CGAP_Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-Sport6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: 0190 dt.
Average insert size 1.7 Kb. Life Technologies catalog #:
11531-019"
BASE COUNT 122 a 78 c 89 g 130 t
ORIGIN
Query Match 13.9%; Score 283; DB 10; Length 419;
Best Local Similarity 99.7%; Pred. No. 2e-78;
Matches 333; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1684 TCCCTGTCCACATCTTCCCATGGGAGATGATCTTCCACAAAGAGCTCACAGCATT 1743
DB 419 TCCCTGTCCACATCTTCCCATGGGAGATGATCTTCCACAAAGAGCTCACAGCATT 360
QY 1744 TTCCACAGAGATGCAATTTTGAGCCCTTGAGATTCCTCAGATTCAGAGAGAAAGTG 1803
DB 359 TTCCACAGAGATGCAATTTTGAGCCCTTGAGATTCCTCAGATTCAGAGAGAAAGTG 300
QY 1804 GGAACAAGTGTGATGGCTACTATGAGCTTGACATCAGCCTATCGTAATCAGAAAT 1863
DB 299 GGAACAAGTGTGATGGCTACTATGAGCTTGACATCAGCCTATCGTAATCAGAAAT 240
QY 1864 ATGAACAATAATCTCTGACACAAAAGCAAGCTTTAAGTTCACAGAGGTCCTGGCTGC 1923
DB 239 ATGAACAATAATCTCTGACACAAAAGCAAGCTTTAAGTTCACAGAGGTCCTGGCTGC 180
QY 1924 ATTGAATATCAGCTCCCTCTGCAATTTCCATCAGATGAGAAGCTTGAGAG 1983
DB 179 ATTGAATATCAGCTCCCTCTGCAATTTCCATCAGATGAGAAGCTTGAGAG 120
QY 1984 CTGCCATCTGTAAATTAATAATTCCTCAATTAAG 2017
DB 119 CTGCCATCTGTAAATTAATAATTCCTCAATTAAG 86
RESULT 10
AM572510/c 417 bp mRNA linear EST 13-MAR-2000
LOCUS xq18g11.x2 NCI-CGAP_Ot1 Homo sapiens cDNA clone IMAGE:2751044 3'
DEFINITION mRNA sequence.

ACCESSION AM572510
 VERSION AM572510.1 GI:7237243
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 417)
 NCICGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCICGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov/image/html/resources.shtml
 Seq primer: -400P from GIBCO
 High quality sequence stop: 416.
 Location/Qualifiers
 1..417
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2751044"
 /clone_lib="NCICGAP-Ut1"
 /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"
 BASE COUNT 140 a 65 c 80 g 132 t
 ORIGIN
 Query Match 7.4%; Score 150; DB 10; Length 417;
 Best Local Similarity 100.0%; Pred. No. 5.1e-37;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1868 AACAAATCTCTGCACAAGAGCAGCTTAAAGTTCACAGSGTGCCTGGCTGCATT 1927
 DB 383 AACAAATCTCTGCACAAGAGCAGCTTAAAGTTCACAGSGTGCCTGGCTGCATT 324
 QY 1928 GAATATCACTTCCCTCTGCATTTTCCCATCACAATAGAACTTGAAGCTGCG 1967
 DB 323 GAATATCACTTCCCTCTGCATTTTCCCATCACAATAGAACTTGAAGCTGCG 264
 QY 1988 CATCTGTAATACATAAATTCACCAATAG 2017
 DB 263 CATCTGTAATACATAAATTCACCAATAG 234
 RESULT 11
 BF056840/c 358 bp mRNA linear EST 16-OCT-2000
 LOCUS 7x10d12.x1 NCICGAP-GC6 Homo sapiens cDNA clone IMAGE:3443783.3
 DEFINITION mRNA sequence.
 ACCESSION BF056840
 VERSION BF056840.1 GI:10810736
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 358)
 NCICGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCICGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.llnl.gov
 Seq primer: -400P from GIBCO.
 FEATURES source
 Location/Qualifiers
 1..358
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3443783"
 /clone_lib="NCICGAP-GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCICGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1459064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 107 a 68 c 82 g 101 t
 ORIGIN
 Query Match 6.7%; Score 136; DB 12; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.3e-32;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1730 AGCTCACAGCATTTTCCACAGATGATGCAATTCGACCCCTTGAGTCCAGTGGATT 1789
 DB 358 AGCTCACAGCATTTTCCACAGATGATGCAATTCGACCCCTTGAGTCCAGTGGATT 299
 QY 1790 CANGAAGAGAGTGGACACAGGTGAGTCTTATGAGCTTGACCATCAGCTAT 1849
 DB 298 CANGAAGAGAGTGGACACAGGTGAGTCTTATGAGCTTGACCATCAGCTAT 239
 QY 1850 CGGTAATCAGAAATAT 1865
 DB 238 CGGTAATCAGAAATAT 223
 RESULT 12
 BF197521/c 525 bp mRNA linear EST 03-NOV-2000
 LOCUS 7084a08.x1 NCICGAP_K1d1 Homo sapiens cDNA clone IMAGE:3642903.3
 DEFINITION similar to TR:Q90BY3 Q90BY3 N-ACETYLDEUCOSANINE 6-O-SULFOTRANSFERASE.; mRNA sequence.
 ACCESSION BF197521
 VERSION BF197521.1 GI:11086670
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 525)
 NCICGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov

FEATURES

High quality sequence stop: 451.
Location/Qualifiers

1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3642903"
/clone_lib="NCI-CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 90 a 163 c 187 g 85 t

ORIGIN

Query Match 2.9%; Score 59; DB 12; Length 525;
Best Local Similarity 100.0%; Pred. No. 9.1e-09;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCGTCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 736
|||||
Db 260 AGGCTGCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 202

RESULT 13
AI824100/c 620 bp mRNA linear EST 21-DEC-1999
LOCUS w146col.x1 NCI-CGAP.Lu19 Homo sapiens CDNA clone IMAGE:2405856 3'
DEFINITION similar to TR:075667 075667 D7J116.4; contains PTRS.b2 PTR5
repetitive element; mRNA sequence.

ACCESSION AI824100 GI:5444771
VERSION AI824100.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 620)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps_email.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Insert Length: 917 Std Error: 0.00
Seq primer: -400P from G1bco
High quality sequence stop: 490.
Location/Qualifiers

FEATURES

1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2405856"
/clone_lib="NCI-CGAP.Lu19"
/tissue_type="squamous cell carcinoma, poorly

differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 106 a 195 c 227 g 89 t 3 others

ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 620;
Best Local Similarity 100.0%; Pred. No. 8e-09;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCGTCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 736
|||||
Db 314 AGGCTGCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGGCTTCTTCAACCTGCAG 256

RESULT 14
BE857538/c 695 bp mRNA linear EST 29-SEP-2000
LOCUS 7901a08.x1 NCI-CGAP.Br23 Homo sapiens CDNA clone IMAGE:3305174 3'
DEFINITION similar to TR:09UBY3 09UBY3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION BE857538 GI:10371664
VERSION BE857538
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 695)
AUTHORS NCI/NIHNS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/STGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps_email.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400P from G1bco
High quality sequence stop: 431.
Location/Qualifiers

FEATURES

1..695
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3305174"
/clone_lib="NCI-CGAP.Br23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTACCAATCTGAAGTGGAGGCGCGCATATCTTTTCTTTTCTTTT
T 3'); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and Eco RI
library is normalized, and was constructed by Bento

BASE COUNT 127 a 210 c 247 g 111 t
 ORIGIN
 Query Match 2.9%; Score 59; DB 12; Length 695;
 Best Local Similarity 100.0%; Pred. No. 7.4e-09;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTTCACCTGCAG 736
 |||||||
 Db 272 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTTCACCTGCAG 214

RESULT 15
 BE858652/c 735 bp mRNA linear EST 29-SEP-2000
 LOCUS 7901a09.x1 NCI-CGAP-Brn23 Homo sapiens cDNA clone IMAGE:3305176 3'
 DEFINITION Similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE ;, mRNA sequence.
 ACCESSION BE858652
 VERSION BE858652.1 GI:10373890
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 735)
 NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Neurological
 Disorders and Stroke, Brain Tumor Genome Anatomy Project
 (CGAP/BTGP), Tumor Gene Index
 UNPUBLISHED (1998)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapsb-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
 Ph.D. Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bernaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 425.
 Location/Qualifiers
 1..735
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI-CGAP-Brn23"
 /tissue_type="glioblastoma (pooled)"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pTZ19-D-Pac (Pharmacia) with a
 modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCTGAAGTGGAGCGGCCGACATCTTTTCTTTTCTTTTCTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pTZ19 vector.
 Library is normalized, and was constructed by Bento
 Soares and M. Fatima Bernaldo."

BASE COUNT 135 a 221 c 253 g 120 t 6 others
 ORIGIN
 Query Match 2.9%; Score 59; DB 12; Length 735;
 Best Local Similarity 100.0%; Pred. No. 7.1e-09;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTTCACCTGCAG 736
 |||||||
 Db 328 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTTCACCTGCAG 270

RESULT 16
 AM081348/c 436 bp mRNA linear EST 14-OCT-1999
 LOCUS xc41b06.x1 NCI-CGAP Co20 Homo sapiens cDNA clone IMAGE:2586803 3'
 DEFINITION similar to TR:088195 088199 CHONDROIN 6-SULFOTRANSFERASE ;, mRNA
 sequence.
 ACCESSION AM081348
 VERSION AM081348.1 GI:6036500
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 436)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 UNPUBLISHED (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgapsb-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
 Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
 I.M.A.G.E. Consortium DNA Sequencing by: Washington University
 Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/btgp/image/image.html
 Seq primer: -400P from Gibco
 High quality sequence stop: 431.
 Location/Qualifiers
 1..436
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI-CGAP-Co20"
 /tissue_type="moderately differentiated adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: SalI;
 Site:2: NotI; Cloned unidirectionally. Primer: oligo dT.
 Normalized to cot 500. Average insert size 1.1kb.
 Normalized version of NCI-CGAP-Co18. Library constructed
 by Life Technologies."

BASE COUNT 64 a 136 c 169 g 67 t
 ORIGIN
 Query Match 2.3%; Score 47; DB 10; Length 436;
 Best Local Similarity 100.0%; Pred. No. 5.7e-05;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTC 724
 |||||||
 Db 309 AGGCTGCGGCTCTACAGCCAGCGTGTGCTCAAGAGAGTCCGCTTC 263

RESULT 17
 BM246681 497 bp mRNA linear EST 01-FEB-2002
 LOCUS K0741E04-3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit-/Sca-1-)
 DEFINITION CDNA Library (Long) Mus musculus cDNA clone K0741E04 3', mRNA
 sequence.
 ACCESSION BM246681
 VERSION BM246681.1 GI:17881951
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 497)
 Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G.,
 Alba, K., Taub, D., Longo, D.D., Keller, J. and Ko, M.S.H.
 Systematic Analyses of NIA Mouse Hematopoietic Stem Cell

JOURNAL
COMMENT

(Lin/-c-Klt-/Sca-1-) cDNA Library (Long)
 Unpublished (2001)
 Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@lgsun.grc.nia.nih.gov
 Plate: K0741 row: E column: 04
 Seq primer: -21M13 Forward
 High quality sequence stop: 497
 PolyA=yes

FEATURES

SOURCE

Location/Qualifiers
 1. 497
 /organism="Mus musculus"
 /strain="C57BL/6NCr"
 /db_xref="taeST:K0741E04-3"
 /db_xref="taxon:10090"
 /clone="K0741E04"
 /clone_lib="NIA Mouse Hematopoietic Stem Cell
 (Lin/-c-Klt-/Sca-1-) cDNA Library (Long)"
 /tissue_type="Hematopoietic Stem Cell (Lin/-c-Klt-/Sca-1-)"
 /dev_stage="Age approx.10 weeks old"
 /lab_host="DH10B"

/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Drs. Dennis Taub, Dan Longo (National Cancer Institute on Aging, USA), Jonathan Keller (National Cancer Institute, USA). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTCTTACATGCGAGCGCCGCTTTT-3'] from 0.9 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-5'. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.1 Kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT
ORIGIN

87 a 137 c 134 g 139 t

Query Match 1.3%; Score 27; DB 13; Length 497;
 Best Local Similarity 100.0%; Pred. No. 87;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 GACATGAGCGTCTTGATGCTACATG 520
 ||||||||||||||||||
 Db 441 GACATGAGCGTCTTGATGCTACATG 467

RESULT 18
BM245312

LOCUS BM245312 536 bp mRNA linear EST 01-FEB-2002
 DEFINITION K0722H07-3 NIA Mouse Hematopoietic Stem Cell (Lin/-c-Klt-/Sca-1-) cDNA Library (Long) Mus musculus cDNA clone K0722H07 3', mRNA sequence.

ACCESSION BM245312
 VERSION BM245312.1 GI:17880582
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS

1 (bases 1 to 536)
 Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Carter,M.G.,
 Alba,K., Taub,D., Longo,D.L., Keller,J. and Ko,M.S.H.
 Systematic Analyses of NIA Mouse Hematopoietic Stem Cell
 (Lin/-c-Klt-/Sca-1-) cDNA Library (Long)
 Unpublished (2001)

TITLE

JOURNAL
COMMENT

Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@lgsun.grc.nia.nih.gov
 Plate: K0722 row: E column: 07
 Seq primer: -21M13 Forward
 High quality sequence stop: 536
 PolyA=yes

FEATURES
SOURCE

Location/Qualifiers
 1. 536
 /organism="Mus musculus"
 /strain="C57BL/6NCr"
 /db_xref="taeST:K0722H07-3"
 /db_xref="taxon:10090"
 /clone="K0722H07"
 /clone_lib="NIA Mouse Hematopoietic Stem Cell
 (Lin/-c-Klt-/Sca-1-) cDNA Library (Long)"
 /tissue_type="Hematopoietic Stem Cell (Lin/-c-Klt-/Sca-1-)"
 /dev_stage="Age approx.10 weeks old"
 /lab_host="DH10B"

/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Drs. Dennis Taub, Dan Longo (National Cancer Institute on Aging, USA), Jonathan Keller (National Cancer Institute, USA). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTCTTACATGCGAGCGCCGCTTTT-3'] from 0.9 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-5'. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.1 Kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT
ORIGIN

93 a 150 c 147 g 146 t

Query Match 1.3%; Score 27; DB 13; Length 536;
 Best Local Similarity 100.0%; Pred. No. 82;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 494 GACATGAGCGTCTTGATGCTACATG 520
 ||||||||||||||||||
 Db 441 GACATGAGCGTCTTGATGCTACATG 467

RESULT 19
BG964671

LOCUS BG964671 695 bp mRNA linear EST 12-JUN-2001
 DEFINITION 602831875F1 NC1_CGAP_Co24 Mus musculus cDNA clone IMAGE:4966258 5', mRNA sequence.

ACCESSION BG964671
 VERSION BG964671.1 GI:14352308
 KEYWORDS EST.
 SOURCE house mouse.

```
ORGANISM      Mus musculus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE         NIH-MGC http://mgc.nci.nih.gov/
JOURNAL       National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT       Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-remail.nih.gov
              Tissue Procurement: Jeffrey E. Green, M.D.
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLM10995 row: f column: 19
              High quality sequence stop: 684.
              Location/Qualifiers
                1..695
                  /organism="Mus musculus"
                  /strain="FVB/N"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:4982643"
                  /clone_1lb="NCI_CGAP_Co24"
                  /lab_host="DH10B (TI phage-resistant)"
                  /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
                  Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                  Average insert size 1.6 kb. Constructed by Life
                  Technologies. Note: this is a NCI_CGAP Library."
              Technology: Note: this is a NCI_CGAP Library."

BASE COUNT    118 a      216 c      195 g      166 t

ORIGIN
Query Match      1.3%: Score 27; DB 13; Length 695;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      494 GACATGAGCGCTTTGATGCTTACATG 520
        |||||||
DB       345 GACATGAGCGCTTTGATGCTTACATG 371

RESULT 20
LOCUS       BG963298              783 bp      mRNA      linear      EST 12-JUN-2001
DEFINITION  60282716F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4982643 5',
ACCESSION   BG963298
VERSION     BG963298
KEYWORDS    EST
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE       NIH-MGC http://mgc.nci.nih.gov/
JOURNAL     National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT     Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-remail.nih.gov
              Tissue Procurement: Jeffrey E. Green, M.D.
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLM10985 row: p column: 04
              High quality sequence stop: 741.
              Location/Qualifiers
                1..783
                  /organism="Mus musculus"
                  /strain="FVB/N"

FEATURES
  source
    Query Match      1.3%: Score 27; DB 13; Length 852;
    Best Local Similarity 100.0%; Pred. No. 58;
    Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    OY      494 GACATGAGCGCTTTGATGCTTACATG 520
            |||||||
    DB       447 GACATGAGCGCTTTGATGCTTACATG 473

RESULT 21
LOCUS       BG966340              952 bp      mRNA      linear      EST 12-JUN-2001
DEFINITION  602832826F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4987327 5',
ACCESSION   BG966340
VERSION     BG966340
KEYWORDS    EST
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE       NIH-MGC http://mgc.nci.nih.gov/
JOURNAL     National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT     Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-remail.nih.gov
              Tissue Procurement: Jeffrey E. Green, M.D.
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLM10998 row: c column: 08
              High quality sequence stop: 784.
              Location/Qualifiers
                1..852
                  /organism="Mus musculus"
                  /strain="FVB/N"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:4987327"
                  /clone_1lb="NCI_CGAP_Co24"
                  /lab_host="DH10B (TI phage-resistant)"
                  /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
                  Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                  Average insert size 1.6 kb. Constructed by Life
                  Technologies. Note: this is a NCI_CGAP Library."
              Technology: Note: this is a NCI_CGAP Library."

BASE COUNT    176 a      248 c      244 g      184 t

ORIGIN
Query Match      1.3%: Score 27; DB 13; Length 852;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      494 GACATGAGCGCTTTGATGCTTACATG 520
        |||||||
DB       447 GACATGAGCGCTTTGATGCTTACATG 473

RESULT 22
LOCUS       AK009113              1000 bp      mRNA      linear      EST 12-JUN-2001
DEFINITION  602832826F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4987327 5',
ACCESSION   AK009113
VERSION     AK009113
KEYWORDS    EST
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE       NIH-MGC http://mgc.nci.nih.gov/
JOURNAL     National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT     Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cga@bbs-remail.nih.gov
              Tissue Procurement: Jeffrey E. Green, M.D.
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLM10998 row: c column: 08
              High quality sequence stop: 784.
              Location/Qualifiers
                1..1000
                  /organism="Mus musculus"
                  /strain="FVB/N"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:4987327"
                  /clone_1lb="NCI_CGAP_Co24"
                  /lab_host="DH10B (TI phage-resistant)"
                  /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
                  Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                  Average insert size 1.6 kb. Constructed by Life
                  Technologies. Note: this is a NCI_CGAP Library."
              Technology: Note: this is a NCI_CGAP Library."

BASE COUNT    176 a      248 c      244 g      184 t

ORIGIN
```


LOCUS AK009113 1923 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003618:carbohydrate (chondroitin 6/keratan) sulfotransferase 4, full insert sequence.
ACCESSION AK009113
VERSION AK009113.1 GI:12843701
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (Strain:C57BL/6J) adult male tongue cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library
clone:2310003618.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
REFERENCE
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,H., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3 Shibata,K., Itoh,M., Aizawa,K., Nagao,K., Sasaki,N., Carninci,P., Kono,H., Akiyama,U., Nishi,K., Katsunai,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishimi,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,K., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system -384-format
sequencing pipeline with 384 multipillar sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Kono,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamakawa,I., Saito,T., Okazaki,Y., Gojohori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pasole,G., Quackenbush,J., Schriml,L.M., Standif,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Balderelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hochmann,M., Hume,D.A., Kamita,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Momperts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Scorch,K.F., Suzuki,H., Togo-Oka,K., Wang,K.H., Weitz,C., Wiltaker,C., Wilming,L., Wyshak-Boris,A., Yoshida,K., Hasegawa,Y., Kawai,H., Kohsaki,S., and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
11217851
5 (bases 1 to 1923)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Balderelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagata,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,

TITLE
JOURNAL
COMMENT
FEATURES
source
Location/Qualifiers
1. 1923
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2310003618"
/db_xref="MGI:190458"
/db_xref="taxon:10090"
/clone="2310003618"
/sex="male"
/tissue.type="tongue"
/clone.lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1. 1923
/gene="Cst4"
96. 1262
/gene="Cst4"
/note="carbohydrate (chondroitin 6/keratan) sulfotransferase 4
data source:MGI, source key:MGI:1349479, evidence:ISS putative"
/codon_start=1
/protein_id="Bab26078.1"
/db_xref="GI:12843702"
/db_xref="MGI:1349479"
/translation="MMLKKGRIMLGSQVIVAFIMSYVHRHISQEESERPVHYLVLSNRSRSGSIVGQLEGGHPDVFYLMERAWHWMTFTSTMKLHVARLDLRVFLCDMSVDAVNWPNPKRQSSIFQWQSRALCSAVCDFPAHEISRKHLKCGQPFDMVEKCRSHGPFVLEKVRFLSLQALYPLLPDSLVLHWVLRDPRVAFVRSREHTIELMVDSHIVYGVHLETKEDQPYVAMKTIKCSQDVIVAIOTLPALQOQVFLFAVEELVRAPIAQTFLKRVFGDLPFLHLOTWYVNTVRGSGOAHFTNARVALVNSQMRWSLPYKVSQLDQACGEMDLIGYLGIVRSQOQEGNGNSLDLSSHLGQVFRNG"
BASE COUNT 445 a 533 c 492 g 453 t
ORIGIN
Query Match 1.3%; Score 27; DB 11; Length 1923;
Best local Similarity 100.0%; Pred. No. 3;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 494 GACATGAGCGCTTTGATGCCCTACATG 520
DB 405 GACATGAGCGCTTTGATGCCCTACATG 431

RESULT 23
 BM036473/c 501 bp mRNA linear EST 05-NOV-2001
 LOCUS fu1f07.x1 Gong zebrafish testis Danio rerio cDNA clone 5377309 3',
 DEFINITION mRNA sequence.
 ACCESSION BM036473
 VERSION BM036473.1 GI:16750044
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 ; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 501)
 AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
 S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
 K., Stepec,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
 Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
 Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
 and Wilson,R.
 TITLE WASHU Zebrafish EST Project 1998
 JOURNAL Unpublished (1998)
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zdraflsh@wustl.edu
 The library was constructed by Dr. Z. Gong. DNA sequencing by:
 Washington University Genome Sequencing Center St. Louis. Please
 contact Zhiyuan Gong for further information on this library
 (National University of Singapore, Department of Biological
 Sciences, Lower Kent Ridge Road, Singapore 119260).
 High quality sequence stop: 367.
 FEATURES
 Location/Qualifiers
 1..501
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone="5377309"
 /clone_lib="Gong zebrafish testis"
 /sex="male"
 /dev_stage="4-5 month"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: testis (pooled); Vector: pBluescript SK-;
 Site 1: XhoI; Site 2: EcoRI; Poly A+ RNA was isolated from
 the testes of 31 male adult zebrafish (4-5 month old)
 cDNAs were made using oligo-dT primers and inserted into
 lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo
 mass-excised to pBluescript SK- following the Washington
 University protocol
 (http://genome.wustl.edu/est/lambda_protocol.shtml).
 Please contact Zhiyuan Gong for further information on
 this library (National University of Singapore,
 Department of Biological Sciences, Lower Kent Ridge Road,
 Singapore 119260).
 BASE COUNT 171 a 94 c 90 g 146 t
 ORIGIN
 Query Match 1.2%; Score 25; DB 13; Length 501;
 Best Local Similarity 100.0%; Pred. No. 3 6e+02;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2008 CCCCAATAGAGAAAAA 2032
 Db 29 CCCCAATAGAGAAAAA 5
 RESULT 24.
 CNS02BKQ 682 bp DNA linear GSS 12-MAY-2000
 LOCUS CNS02BKQ
 DEFINITION Tetradon nigroviridis genome survey sequence PCC-0r1 end of clone
 253N23 of library G from Tetradon nigroviridis, genomic survey
 sequence.

ACCESSION AL189971
 VERSION AL189971.1 GI:7828075
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 REFERENCE 1 (bases 1 to 682)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 TITLE Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 682)
 AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis
 JOURNAL Submitted (12-APR-2000)
 COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetradon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 Location/Qualifiers
 1..682
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="253N23"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0AG253CG12SP1-end ;
 PCC-0r1"
 BASE COUNT 202 a 122 c 113 g 224 t 21 others
 ORIGIN
 Query Match 1.2%; Score 25; DB 17; Length 682;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2008 CCCCAATAGAGAAAAA 2032
 Db 241 CCCCAATAGAGAAAAA 217
 RESULT 25
 BM663231/c 375 bp mRNA linear EST 27-FEB-2002
 LOCUS BM663231
 DEFINITION UI-E-C10-aah-e-08-0-01.s1 UI-E-C10 Homo sapiens cDNA clone
 BM663231
 ACCESSION BM663231.1 GI:18968394
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 375)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrived by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 The following repetitive elements were found in this CDNA sequence: 1-42, >POLY_A#Simple_repeat (matched compliment)
 Seq primer: M13 Forward
 POLYA=yes

FEATURES

SOURCE

Location/Qualifiers
 1..375
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-C10-aah-e-08-0-UI"
 /clone_lib="UI-E-C10"
 /tissue_type="RPE and Choroid"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I;
 UI-E-C10 is a CDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
 TAG_LIB=UI-E-C10
 TAG_TISSUE=RPE and Choroid
 TAG_SEQ=ACCTA"

BASE COUNT 88 a 82 c 89 g 116 t
 ORIGIN

Query Match 1.2%; Score 24; DB 13; Length 375;
 Best Local Similarity 100.0%; Pred. No. 9.2e+02;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2003 CCAATATAGAAAAA 2032
 Db 26 CCAATATAGAAAAA 3

RESULT 26
 BM666130/c 566 bp mRNA linear EST 27-FEB-2002

LOCUS UI-E-DX1-agt-f-19-0-UI.s1 UI-E-DX1 Homo sapiens CDNA clone
 DEFINITION UI-E-DX1-agt-f-19-0-UI 3', mRNA sequence.

ACCESSION BM666130
 VERSION BM666130.1 GI:18973587
 KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 566)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477
 CONTACT: Soares, MB
 PROGRAM FOR Rat Gene Discovery and Mapping
 UNIVERSITY OF IOWA

FEATURES

SOURCE

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrived by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 The following repetitive elements were found in this CDNA sequence: 23-556, >11#LINE/L1 (matched compliment)
 Seq primer: M13 Forward
 POLYA=yes

Location/Qualifiers
 1..566
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-DX1-agt-f-19-0-UI"
 /clone_lib="UI-E-DX1"
 /tissue_type="fetal eyes"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I;
 UI-E-DX1 is a normalized CDNA library containing the following tissue(s): fetal eyes. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACATCAAGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
 TAG_LIB=UI-E-DX1
 TAG_TISSUE=human fetal eyes
 TAG_SEQ=ACATCAAGA"

BASE COUNT 108 a 94 c 89 g 275 t
 ORIGIN

Query Match 1.2%; Score 24; DB 13; Length 566;
 Best Local Similarity 100.0%; Pred. No. 6.8e+02;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATATAGAAAAA 2032
 Db 24 CCAATATAGAAAAA 1

RESULT 27

CNS0127Y 919 bp DNA linear GSS 12-MAY-2000

LOCUS Tetraodon nigroviridis genome survey sequence pUC-ori end of clone
 DEFINITION 219J08 of library G from Tetraodon nigroviridis, genomic survey

ACCESSION AL173959
 VERSION AL173959.1 GI:7812016

KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 919)
 Roest-Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.

```

TITLE      Human gene number estimate provided by genome wide analysis using
JOURNAL    Tetradon nigroviridis DNA sequence
REFERENCE   Unpublished
AUTHORS     2 (bases 1 to 919)
            Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bounneau-L., Billault,A., Querier,F., Saurin,W., Bernot,A. and
            Weissensbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
JOURNAL    freshwater pufferfish Tetradon nigroviridis
REFERENCE   Unpublished
AUTHORS     3 (bases 1 to 919)
JOURNAL    Direct Submission
TITLE      Genoscope.
COMMENT     Submitted (12-APR-2000)
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetradon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/tetradon.
FEATURES    Location/Qualifiers
            1..919
            /organism="Tetradon nigroviridis"
            /db_xref="taxon:99863"
            /clone="219308"
            /clone_1lb="G"
            /note="Genoscope sequence ID : COAG219DE04SP1-end :
            PUC-ori"
BASE COUNT  314 a    151 c    164 g    283 t    7 others
ORIGIN
Query Match      1.2%; Score 24; DB 17; Length 919;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  2009 CCAATTAGAGAAAAA 2032
Db   514 CCAATTAGAGAAAAA 537

RESULT 28
B75634/c      132 bp    DNA    linear    GSS 08-APR-1999
LOCUS        RPCI11-11M14.TV RPCI-11 Homo sapiens genomic clone RPCI-11-11M14,
DEFINITION   DNA sequence.
ACCESSION    B75634
VERSION      B75634.1 GI:2771321
KEYWORDS     GSS.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 132)
            Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
            ,J.C.
            Use of BAC End Sequences for Sequence-Ready Map Building
JOURNAL      Unpublished (1997)
COMMENT      Other-GSS: RPCI11-11M14.TP
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@jg.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES     Location/Qualifiers

```

```

source      1..132
            /organism="Homo sapiens"
            /db_xref="GDB:7504141"
            /db_xref="taxon:9606"
            /clone="RPCI-11-11M14"
            /clone_1lb="RPCI-11"
            /sex="Male"
            /cell_type="Lymphocytes"
            /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
            RPCI11 Human Male BAC Library"
BASE COUNT  36 a    19 c    20 g    57 t
ORIGIN
Query Match      1.1%; Score 23; DB 17; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.2e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  2010 CAAATAGAAAAA 2032
Db   107 CAAATAGAAAAA 85

RESULT 29
A1072478/c     191 bp    mRNA    linear    EST 05-JUL-1999
LOCUS        UI-R-C2-nm-c-03-0-UI.s1 UI-R-C2 Rattus norvegicus cDNA clone
DEFINITION   UI-R-C2-nm-c-03-0-UI 3', mRNA sequence.
ACCESSION    A1072478
VERSION      A1072478.1 GI:3398672
KEYWORDS     EST.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 191)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
COMMENT      MEDLINE
            97044477
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@iue.uiowa.edu
            cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
            distribution: clones will be available through Research Genetics
            This clone is also available through the I.M.A.G.E. Consortium at
            LNL (info@image.llnl.gov). IMAGE ID=1786676
            Seq primer: M13 Forward
            POLYA=No.
FEATURES     Location/Qualifiers
            1..191
            /organism="Rattus norvegicus"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /clone="UI-R-C2-nm-c-03-0-UI"
            /clone_1lb="UI-R-C2"
            /dev_stage="adult"
            /lab_host="DH10B (Life Technologies)"
            /note="Vector: pRT3D-Pac (Pharmacia) with a modified
            polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-C2
            library is a subtracted library derived from the UI-R-C1
            library, which is a subtracted library derived from the
            UI-R-C0 library. The UI-R-C0 library consisted of a
            mixture of individually tagged normalized libraries
            constructed from rat placenta, adult lung, brain, liver,
            kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
            embryo. The tag is a string of 3-5 nucleotides present
            between the Not I site and the oligo-dT track which allows

```

BASE COUNT 47 a 41 c 30 g 73 t
ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 191;
Best Local Similarity 100.0%; Pred. No. 3.1e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAGAGAAAAA 2032
Db 25 CAATAGAGAAAAA 3

RESULT 30 269 bp mRNA linear EST 28-JUN-2000
B0128285 RIKEN full-length enriched, 16 days neonate cerebellum Mus
LOCUS B0128285
DEFINITION Musculus cDNA clone 9630029B21 3', mRNA sequence.
ACCESSION B0128285
VERSION B0128285.1 GI:8782652
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE

AUTHORS Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
1 (bases 1 to 269)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arkawa, T., Carninci
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, T., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Konno, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata
Y., Shigemoto, Y., Shinagawa, A., Shirai, T., Sobabe, Y., Suganara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takehashi, F., Tomiyaga, N., Toya
T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamakata, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC) Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
Carninci, P. and Hayashizaki, Y.
URL: http://genome.gsc.riken.go.jp/

TITLE

JOURNAL

COMMENT

Identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2 library. This procedure has been previously described (Bonaldi, Lennon and Soares, Genome Research 6: 791-806, 1996).

Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (3), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

FEATURES

source

19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers

1..269

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="9630029B21"

/clone_11b="RIKEN full-length enriched, 16 days neonate cerebellum"

/tissue_type="cerebellum"

/dev_stage="16 days neonate"

/lab_host="DH10B"

/note="Site 1: SalI; Site 2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',

GAGACAGAGAGATCCAGACCTCTTTTCTTTTCTTTT 3']. cDNA was prepared by using trianose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTGAGTATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified phuescript KS(+) after bulk excision from Lambda FIC I."

BASE COUNT 79 a 52 c 30 g 108 t
ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 269;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATAGAGAAAAA 2032
Db 159 CAATAGAGAAAAA 137

RESULT 31 304 bp mRNA linear EST 30-APR-2002
B0191911/c
LOCUS B0191911
DEFINITION UI-R-DRI-CKY-g-20-0-UI-s1 UI-R-DRI Rattus norvegicus cDNA clone
UI-R-DRI-CKY-g-20-0-UI s1, mRNA sequence.
ACCESSION B0191911
VERSION B0191911.1 GI:20367462
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE

AUTHORS Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae:
Rattus.
1 (bases 1 to 304)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

Genome Res. 6 (9), 791-806 (1996)

TITLE

JOURNAL

COMMENT

Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (3), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the

normalized osteoblast library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLA=yes

FEATURES

Source

Location/Qualifiers

1. 304
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DRI-cky-g-20-0-UI"
/clone_lib="UI-R-DRI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-DRI
library is a normalized Rat Osteoblast library (nREO)
constructed in pT73 vector according to the procedure
described by Bonaldo, Lennon & Soares (Normalization and
Subtraction: Two Approaches to Facilitate Gene Discovery.
Genome Research 6: 791-806, 1996). The oligonucleotide
used to prime first strand synthesis contained the
sequence tag AACATATCAA between the Not I cloning site and
dri8 stretch. The Rat Osteoblast tissue was provided by
Lian & Stein of the University of Massachusetts Medical
School.

TAG_LTB=UI-R-DRI
TAG_TISSUE=osteoblast
TAG_SEQ=AACATATCAA"

BASE COUNT 84 a 67 c 49 g 104 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CCAATAGAAAAA 2032

Db 23 CCAATAGAAAAA 1

RESULT 32
BE117198/c 319 bp mRNA linear EST 13-JUN-2000
LOCUS BE117198
DEFINITION UI-R-BS1-ayh-a-02-0-UI.s1 UI-R-BS1 Rattus norvegicus cDNA clone
UI-R-BS1-ayh-a-02-0-UI 3', mRNA sequence.
BE117198
VERSION BE117198.1 GI:8509303
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 319)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)

REFERENCE 1
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msocares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonalde poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
The following repetitive elements were found in this cDNA sequence:

1-48, >POLY_A#Simple_repeat
Seq primer: M13 Forward
POLA=yes

FEATURES

Source

Location/Qualifiers

1. 319
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BS1-ayh-a-02-0-UI"
/clone_lib="UI-R-BS1"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BS1
library is derived from 13 dpc whole embryo tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
rategen.eng.uiowa.edu.
TAG_SEQ=None found"

BASE COUNT 67 a 63 c 61 g 128 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATAGAAAAA 2031

Db 23 CCAATAGAAAAA 1

RESULT 33
BG291580 342 bp mRNA linear EST 21-FEB-2001
LOCUS BG291580
DEFINITION 60238567P1 NIH-MGC_93 Homo sapiens cDNA clone IMAGE:4514872 5',
mRNA sequence.
BG291580
VERSION BG291580.1 GI:13049825
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 342)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM10403 row: m column: 17
High quality sequence stop: 342.

FEATURES

Source

Location/Qualifiers

1. 342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4514872"
/clone_lib="NIH-MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: pCMV-SPORT6. Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."

BASE COUNT 136 a 49 c 59 g 98 t
ORIGIN

Query Match 1.1%; Score 23; DB 12; Length 342;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032
 ||||||||||||||||
 Db 294 CAATATAGAAAAA 316

RESULT 34
 B1555949 349 bp mRNA linear EST 05-SEP-2001
 LOCUS 603239076F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5291971 5',
 DEFINITION mRNA sequence.
 ACCESSION B1555949 GI:15443263
 VERSION B1555949
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 349)
 NIH-MGC http://mgi.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Chu Xia Deng Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1738 row: h column: 20
 High quality sequence stop: 349.

FEATURES
 source
 Location/Qualifiers
 1..349
 /organism="Mus musculus"
 /strain="129,C57BL/6J,FVB/N"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:5291971"
 /clone_id="NCI_CGAP_Mam3"
 /tissue_type="tumor, gross tissue"
 /dev_stage="10 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; Salt:
 Site:2; Notti: Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigators
 Providing samples: Lothar Hennighausen/Chu Xia Deng, NIH
 Reference for transgenic model: Xu et al., Nature Genetics
 22, 37-43 (1999)."

BASE COUNT 108 a 80 c 66 g 95 t
 ORIGIN

Query Match 1.1%; Score 23; DB 13; Length 349;
 Best Local Similarity 100.0%; Pred. No. 2e+03;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAATATAGAAAAA 2032
 ||||||||||||||||
 Db 327 CAATATAGAAAAA 349

RESULT 35
 BE340900 362 bp mRNA linear EST 17-JUL-2000
 LOCUS EST344988 potato stolon, Cornell University Solanum tuberosum cDNA
 DEFINITION clone cSTAL1K13, mRNA sequence.
 ACCESSION BE340900 GI:9250431
 VERSION BE340900.1
 KEYWORDS EST.

SOURCE potato.
 ORGANISM Solanum tuberosum

REFERENCE 1 (bases 1 to 362)
 van der Hoeven, R., Bezzerides, J., Bachem, C., Horvath, B., Visser, R.,
 Holt, I.E., Liang, F., Hansen, T.S., Uiterback, T., Bowman, C.L., Doan
 'B., Bougri, O., Buell, C.R., Ronning, C.M., Tanksley, S.D. and Baker
 'B.
 Generation of ESTs from potato swelling stolons

JOURNAL Unpublished (1999)
 COMMENT Contact: Research Genetics, Libraries Division
 Tel: 1-800-711-6195
 Email: cdna@resgen.com
 5 prime sequence.
 FEATURES
 source
 Location/Qualifiers
 1..362
 /organism="Solanum tuberosum"
 /cultivar="Bintje"
 /db_xref="taxon:4113"
 /clone_lib="cSTAL1K13"
 /clone_id="potato stolon, Cornell University"
 /tissue_type="axillary buds of stem explants, swelling
 stolons"
 /dev_stage="1 to 3 days"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site:1: EcoRI; Site:2:
 XhoI; RNA was supplied by Christian Bachem & Beatrix
 Horvath/Laboratory of Plant Breeding, Dept. of Plant
 Sciences, Wageningen University, The Netherlands). Total
 RNA was isolated from developing axillary buds of potato
 nodal stem cuttings cultured on medium for the
 introduction of tuber formation as described in Bachem et
 al. (Plant Journal 1996). Tissue samples were taken of
 stages corresponding to growing stolons and the early
 stages of tuber formation."

BASE COUNT 121 a 60 c 74 g 107 t
 ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 362;
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1575 ATTGCTCTCTCTCTCTCTGAT 1597
 ||||||||||||||||
 Db 58 ATTGCTCTCTCTCTCTCTGAT 36

RESULT 36
 BM930440 385 bp mRNA linear EST 13-MAR-2002
 LOCUS BM930440

DEFINITION UI-E-EJ1-8jm-c-11-0-UI-1 UI-E-EJ1 Homo sapiens cDNA clone
 UI-E-EJ1-8jm-c-11-0-UI 5', mRNA sequence.

ACCESSION BM930440
 VERSION BM930440.1 GI:19389613
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 385)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: msouresblue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 REVERSE.

FEATURES

Source
 Location/Qualifiers
 1..385
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="U1-E-EJ1-a1m-c-11-0-01"
 /clone_lib="U1-E-EJ1"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina Foveal and Macular, RPE and
 choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 U1-E-EJ1 is a subtracted cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes, AGAATCAGA
 ; lens, CGATTACGGA; eye anterior segment, AATGCCGAT;
 optic nerve, CCAATAGAG; retina, CCGCG; Retina Foveal and
 Macular, GTCC; RPE and Choroid, ACCGA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."
 BASE COUNT 156 a 41 c 99 g 89 t
 ORIGIN

Query Match 1.1%; Score 23; DB 14; Length 385;
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2010 CAAATAGAGAAAAA 2032
 ||||||||||||||||||
 Db 356 CAAATAGAGAAAAA 378

RESULT 37
 B1505360 397 bp mRNA linear EST 08-APR-2002
 LOCUS B1170026A10E04.5 Bee Brain Normalized/Subtracted Library, Bb17 Apis
 DEFINITION mellifera cDNA clone Bb170026A10E04 5', mRNA sequence.
 ACCESSION B1505360
 VERSION B1505360.1 GI:15355734
 KEYWORDS EST.
 SOURCE honeybee.
 ORGANISM Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prexygota;
 Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
 Apidae; Apis.
 1 (bases 1 to 397)
 Whitefield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
 Pardinas,T., Robertson,H.M., Soares,B. and Robinson,G.E.
 Annotated expressed sequence tags and cDNA microarrays for studies
 of brain and behavior in the honey bee
 Genome Res. 12 (4), 553-566 (2002)
 JOURNAL 21929762
 MEDLINE
 COMMENT Contact: Gene E. Robinson
 Department of Entomology
 University of Illinois
 505 S. Goodwin Ave., Urbana, IL 61801, USA
 Tel: 217 265 0305

Fax: 217 244 3499
 Email: generob@life.uiuc.edu
 This research was funded by the University of Illinois Critical
 Research Initiatives fund and a Burroughs-Wellcome Trust Innovation
 Award in Functional Genomics to G.E. Robinson and an NSF
 Postdoctoral Fellowship in Bioinformatics to C.W. Whitefield.
 REPEATS IN THE SEQUENCE
 Low complexity STRAND (+) ELEMENT AT rich LOCATION [220,279];
 Simple Repeat STRAND (+) ELEMENT (TTAA)n LOCATION [281,354];
 Simple Repeat STRAND (+) ELEMENT (A)n LOCATION [357,393].
 PCR Primers
 FORWARD: TAAATGACCTCCTATAGG
 BACKWARD: ATTACCTCCTATAGG
 Plate: Bb170026A10 row: E column: 04
 Seq primer: AGCGATACCAATTTACACAGGA
 High quality sequence stop: 397.
 FEATURES
 Source
 Location/Qualifiers
 1..397
 /organism="Apis mellifera"
 /strain="mixed strains of European bees, predominantly
 A.m. ligustica"
 /db_xref="taxon:7460"
 /clone="Bb170026A10E04"
 /clone_lib="Bee Brain Normalized/Subtracted Library, Bb17"
 /sex="female"
 /tissue_type="brain"
 /dev_stage="adult worker honey bee"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pT73-Pac; Site_1: EcoRI;
 Site_2: NotI. This Bb17 cDNA library was generated by
 subtraction of the Bb16 library with 4000 previously
 sequenced clones. The Bb16 library was contributed by the
 Soares laboratory and it was constructed and normalized
 as described by Bonaldo, M.F., Lennon, G. and Soares,
 M.B. (1996), Genome Research 6(9): 791-806. RNA was
 prepared from dissected brains of adult worker bees of
 various ages and various behavioral groups."

BASE COUNT 175 a 47 c 50 g 123 t
 ORIGIN

Query Match 1.1%; Score 23; DB 13; Length 397;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2010 CAAATAGAGAAAAA 2032
 ||||||||||||||||||
 Db 360 CAAATAGAGAAAAA 382

RESULT 38
 BE109147 408 bp mRNA linear EST 13-JUN-2000
 LOCUS BE109147
 DEFINITION U1-R-BS1-azb-e-02-0-UI.s1 U1-R-BS1 Rattus norvegicus cDNA clone
 ACCESSION BE109147
 VERSION BE109147.1 GI:8501252
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 408)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250

Fax: 319 335 9565
Email: mscares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized embryo at 13 dpc library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-48,
>POLY_ASimple_repeat
Seq primer: M13 Forward
POLYA=yes

FEATURES

Source

Location/Qualifiers
1..408
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UT-R-BS1-azd-e-02-0-UT"
/clone_1lb="UT-R-BS1"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/note="Vector: PT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The UT-R-BS1 library is derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.cmg.uiowa.edu.
TAG_L1b=UT-R-BS1
TAG_TISSUE=embryo at 13 dpc
TAG_SEQ=AATCC"
BASE COUNT 73 a 102 c 76 g 157 t
ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 408;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATATAGAAAAA 2031
Db 23 CCAATATAGAAAAA 1

RESULT 39
AI095726/c 413 bp mRNA linear EST 05-OCT-1998
LOCUS 922412.x1 Soares-pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION IMAGE:1697231 3', mRNA sequence.
ACCESSION AI095726
VERSION AI095726.1 GI:3434702
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 413)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 1443 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 405.
Location/Qualifiers
1..413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1697231"

/clone_1lb="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: PT73-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15',
AAGCGAAGATTCGGCCCGCCCTTTTCTTTTCTTTT 3',
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified PT73 vector. Library went through one round of normalization. Library constructed by M. Patricia Bonafide."
BASE COUNT 96 a 93 c 105 g 119 t
ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 413;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATATAGAAAAA 2031
Db 23 CCAATATAGAAAAA 1

RESULT 40
AM080577/c 419 bp mRNA linear EST 14-OCT-1999
LOCUS xc33e07.x1 NCI-CGAP-Co18 Homo sapiens cDNA clone IMAGE:2586084 3'
DEFINITION similar to gb:DI3315 LACTOYLGLUTATHIONE LYASE (HUMAN);, mRNA sequence.
ACCESSION AM080577
VERSION AM080577.1 GI:6035729
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 419)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Meskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/bbnp/image/image.html
Seq primer: 40UP from GIBCO.
Location/Qualifiers
1..419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2586084"
/clone_1lb="NCI-CGAP-Co18"
/tissue_type="moderately differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.26 kb. Library constructed by Life Technologies. Normalized versions of this library named NCI-CGAP-Co19 (Cot 50), NCI-CGAP-Co20 (Cot 500), and NCI-CGAP-Co21 (Cot >500)."
BASE COUNT 130 a 72 c 69 g 148 t
ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAGAAAAA 2032
 |||||||
 Db 31 CAAATAGAGAAAAA 9

RESULT 41
 BE107411/c 420 bp mRNA linear EST 13-JUN-2000
 LOCUS UI-R-BS1-ayt-d-07-0-UI-s1 UI-R-BS1 Rattus norvegicus cDNA clone
 DEFINITION UI-R-BS1-ayt-d-07-0-UI 3', mRNA sequence.
 ACCESSION BE107411
 VERSION BE107411.1 GI:8495516
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 420)
 REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.
 AUTHORS Normalization and subtraction: two approaches to facilitate gene
 TITLE discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag served in the cDNA between the NotI site
 and the oligo-dT track present to verify it as a clone from the
 normalized embryo at 13 dpc library cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: clones will be available through
 Research Genetics (www.resgen.com) The following repetitive
 elements were found in this cDNA sequence: 1-48,
 >POLY_ASimple_repeat
 Seq primer: M13 Forward
 POLY_A=Yes.

FEATURES
 source
 Location/Qualifiers
 1..420
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BS1-ayt-d-07-0-UI"
 /clone_lib="UI-R-BS1"
 /dev_stage="embryonic 13 dpc"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BS1
 library is derived from 13 dpc whole embryo tissue. For a
 detailed description of the library from which this clone
 was derived, please visit our web site at
 ratest.eng.uiowa.edu.
 TAG_LIB=UI-R-BS1
 TAG_TISSUE=embryo at 13 dpc
 TAG_SEQ=ATATCC"
 BASE COUNT 76 a 104 c 80 g 160 t
 ORIGIN
 Query Match 1.1%; Score 23; DB 10; Length 420;
 Best Local Similarity 100.0%; Pred. NO. 1.7e+03;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATATAGAGAAAAA 2031
 |||||||
 Db 23 CCAATATAGAGAAAAA 1

RESULT 42
 A1195993/c 423 bp mRNA linear EST 14-OCT-1998
 LOCUS uic3b09.x1 Sugano mouse liver tlia Mus musculus cDNA clone
 DEFINITION IMAGE:1887065.3', similar to gb:M16961 ALPHA-2-HS-GLYCOPROTEIN
 PRECURSOR (HUMAN);, mRNA sequence.
 ACCESSION A1195993
 VERSION A1195993
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 423)
 REFERENCE Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,D., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:971389
 Seq primer: custom primer used
 High quality sequence stop: 409.

FEATURES
 source
 Location/Qualifiers
 1..423
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1887065"
 /clone_lib="Sugano mouse liver tlia"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
 (CACTGATG); Site_2: DraIII (CACCATGTC); 1st strand cDNA
 was primed with an oligo(dT) primer
 [ATGTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor (TTTGGCCTCTGCG), digested
 and cloned into distinct DraIII sites of the pME18S-FL3
 vector (5' site CACTGATG, 3' site CACCATGTC). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTTGTGCTTAAGAAGTGGC and 3' end
 primer CGACCTGCACCTCGAGCAC."
 BASE COUNT 89 a 103 c 106 g 125 t
 ORIGIN
 Query Match 1.1%; Score 23; DB 9; Length 423;
 Best Local Similarity 100.0%; Pred. NO. 1.7e+03;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAGAAAAA 2032
 |||||||
 Db 23 CAAATAGAGAAAAA 1

RESULT 43
 W32300/c 437 bp mRNA linear EST 10-OCT-1996
 LOCUS

DEFINITION zc05905.s1 Soares-Parathyroid-tumor-NBHPA Homo sapiens cDNA clone IMAGE:321464 3' similar to gb:D13315 LACTOYLGLUTATHIONE LYASE (HUMAN); mRNA sequence.

ACCESSION W33300
VERSION W33300.1 GI:1313311
KEYWORDS EST.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 437)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holtan,M., Hulman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rikhi,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
The Washu-Merck EST Project
Unpublished (1995)
CONTACT: Wilson R.
JOURNAL Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1429 Std Error: 0.00
Seq primer: mob, REGA+ET
High quality sequence stop: 374.

FEATURES
Source
Location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="GDB:1259122"
/db_xref="taxon:9606"
/clone="IMAGE:321464"
/clone_lib="Soares-Parathyroid-tumor-NBHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pTZ19 (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer
15'-GTGTACCAATCTGAAGTGGAGCGCGACCAATTTTTTTTTTTTTTTT
TTTTT-3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTZ19 vector (Pharmacia). Library went through one round of normalization to a cot = 5. Library constructed by Bento Soares and M. Fatima Bonaldi. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."

BASE COUNT 137 a 75 c 73 g 151 t 1 others
ORIGIN

Query Match 1.1%; Score 23; DB 14; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2010 CAATATAGAAAAA 2032
|||||
Db 25 CAATATAGAAAAA 3

RESULT 44
AI952475/c 443 bp mRNA linear EST 09-MAR-2000
LOCUS w74c04.x1 NCI CGAP Ov38 Homo sapiens cDNA clone IMAGE:2549382 3'
DEFINITION similar to gb:D13315 LACTOYLGLUTATHIONE LYASE (HUMAN); mRNA
Sequence.
ACCESSION AI952475.
VERSION AI952475.1 GI:5744785
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 443)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Jeff Struwing, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA library preparation: Life Technologies, Inc. cDNA library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html
Insert length: 2122 Std Error: 0.00
Seq primer: -40up from Gibco
High quality sequence stop: 396.

FEATURES
Source
Location/Qualifiers
1..443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2549382"
/clone_lib="NCI-CGAP-Ov38"
/tissue_type="normal epithelium"
/lab_host="DH10B"
/note="Organ: ovary; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

BASE COUNT 139 a 77 c 75 g 152 t
ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 443;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2010 CAATATAGAAAAA 2032
|||||
Db 27 CAATATAGAAAAA 5

RESULT 45
B0013210/c 446 bp mRNA linear EST 26-MAR-2002
LOCUS UT-1-BCLP-ayt-d-07-0-01.s1 NCI-CGAP-P13 Homo sapiens cDNA clone
DEFINITION UT-1-BCLP-ayt-d-07-0-01 3', mRNA sequence.
ACCESSION B0013210
VERSION B0013210.1 GI:19738111
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 446)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Steven Brown
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA-yes.
Location/Qualifiers
1..446

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-1-BC1P-act-d-07-0-UI"
/clone_lib="NCI-CGAP_p13"
/tissue_type="placenta"
/dev_stage="8-9 weeks"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Placenta; Vector: pT73-Pac (Pharmacia) with
a modified polylinker; Site_1: EcoR I; Site_2: Not I;
NCI-CGAP_p13 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are GA, AGCA. For
additional information, contact: Bento Soares,
bento-soares@uiowa.edu
TAG_LIB=UI-1-BC1P
TAG_TISSUE=placenta human 8 week
TAG_SEQ=GA"

BASE COUNT      94 a      119 c      105 g      128 t
ORIGIN

Query Match      1.1%; Score 23; DB 14; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAGAAAAA 2032
|||||
Db 23 CAAATAGAGAAAAA 1

RESULT 46
BQ427521      465 bp      mRNA      linear      EST 24-MAY-2002
LOCUS      AGENCOURL_7919274 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6160346
DEFINITION      5', mRNA sequence.
ACCESSION      BQ427521
VERSION      BQ427521.1 GI:21166597
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 465)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabsr@mail.nih.gov
Tissue procurement: ATCC/DCMD/DPMP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM13510 row: 0 column: 03
High quality sequence stop: 464.
Location/Qualifiers
1. 465
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6160346"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (Phage-resistant)"
/note="Organ: skin; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

```

```

Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT      155 a      83 c      82 g      144 t      1 others
ORIGIN

Query Match      1.1%; Score 23; DB 14; Length 465;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2010 CAAATAGAGAAAAA 2032
|||||
Db 442 CAAATAGAGAAAAA 464

RESULT 47
BQ194968      481 bp      mRNA      linear      EST 30-APR-2002
LOCUS      BQ194968
DEFINITION      UI-R-CNI-cmd-k-19-0-UI.s1 UI-R-CNI Rattus norvegicus cDNA clone
UI-R-CNI-cmd-k-19-0-UI 3', mRNA sequence.
ACCESSION      BQ194968
VERSION      BQ194968.1 GI:20370519
KEYWORDS      EST.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 481)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msocres@iuiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonalde poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized fundus library cDNA library Preparation: M.B. Soares Lab
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers
1. 481
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CNI-cmd-k-19-0-UI"
/clone_lib="UI-R-CNI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CNI
library is a subtracted library derived from the following
pool of seven normalized rat libraries: normalized rat
seminal vesicles, normalized rat penis, normalized rat
bladder, normalized rat cervix, normalized rat brown
adipose, normalized rat fundus, and normalized rat
salivary gland. It was constructed according to the
procedure described by Bonaldo, Lennon & Soares (genome
Research genome 6: 791-806, 1996). For construction of
the CNI library, plasmid DNA from the pool of normalized
libraries was electroporated into competent bacteria for
the production of single-stranded circular DNA. This was
then used as a tracer in a subtractive hybridization with

```

a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat Urogen Set corresponding to plates R-5-AA-NN excluding plates R-5-MM and NM. This pool represented 40% of the final driver population. b) a pool of about 29,000 clones from subcloned libraries C40 and C41 corresponding to plates R-C40-AMV through R-C40-AS, R-C40-AZX through R-C40-BAZ, R-C40-BEE through R-C40-BHY, R-C40-BJS, R-C40-BKE, R-C40-BKG-H, R-C40-BKJ-K, R-C40-BKE through R-C40-BKS, R-C40-BKV-V, R-C40-BLY through R-C40-BMA, R-C40-BMC through R-C40-BME, R-C40-BNS, R-C40-BOB through R-C40-BOJ, R-C40-BPA through R-C40-BPG, R-C41-BBA through R-C41-BDA, R-C41-BHZ through R-C41-BIF, R-C41-BIR, R-C41-BJT through R-C41-BKB, R-C41-BKD, R-C41-BK, R-C41-BKI, R-C41-BKT, R-C41-BLF, R-C41-BLH through R-C41-BLN, R-C41-BLS, R-C41-BLV, R-C41-BMR, and R-C41-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non-normalized libraries CS0s, CT0s, CU0s, CM0s, CX0s and normalized libraries CS0, CT0, CU0, CM0, and CX0 corresponding to plates R-CS0s-CBD through R-CS0s-CBO, R-CT0s-CAM through R-CT0s-CAV, R-CU0s-CBP through R-CU0s-CCB, R-CM0s-CCB through R-CM0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN, R-CU0-BDQ through R-CU0-BVL, R-CM0-BVU through R-CM0-BMP, R-CM0-BXN through R-CM0-BXO, R-CX0-BWQ through R-CX0-BXM. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from non-normalized eye library CVO and 4,000 from normalized eye library CV1) corresponding to plates R-CVO-BSD through R-CVO-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subcloned library BS2, BVO and BV0p (7.9.5 kb cDNA library fraction from rat whole embryo), and BX0 (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0p-NOI through R-BV0p-NOX, and R-BX0-AXQ through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CY0, CZ0, DA0, DB0, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BYA through R-CZ0-BYL, R-DA0-BZB-C, R-DB0-BVJ through R-DB0-BXP, R-DC0-BZD through R-DC0-BZH, R-DD0-BXQ through R-DD0-BZA, R-DD0-BZI through R-DD0-BZQ, R-DD0-CAY through R-DD0-CBA, R-DD0-BZR through R-DD0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CNO, that makes up the tracer. The corresponding plates are R-CNO-BKV through R-CNO-BLD, R-CNO-BIG, R-CNO-BIP through R-CNO-BIR, R-CNO-BLT, R-CNO-BLW-X, R-CNO-BMB, and R-CNO-BMF through R-CNO-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CNO pool corresponding to the following addresses: bkx-a-09-0-UI, bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI, bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI, bkx-a-05-0-UI, bkx-a-06-0-UI, bkx-a-11-0-UI, bkx-z-06-0-UI, bkx-z-08-0-UI, bkx-z-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, bla-h-12-0-UI, bla-h-02-0-UI, bla-c-a-11-0-UI, bla-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, bly-g-04-0-UI, bly-a-05-0-UI, bly-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CNO clone (corresponding to the address bkx-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population.

FEATURES
source

1.484
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0p-afd-c-06-0-UI"
/clone.lib="UI-R-BJ0p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTR3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-BJ0p library is a subcloned library derived from the UI-R-AA1, UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and UI-R-AG1 libraries. These libraries represent tissues from rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, and ventricle at 13 dpc. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research

2010 CAATATAGAAAAA 2032
DB 23 CAATATAGAAAAA 1

RESULT 48
AM434101/c
LOCUS
DEFINITION
UI-R-BJ0p-afd-c-06-0-UI s1 UI-R-BJ0p Rattus norvegicus cDNA clone
VERSION
AM434101.1 GI:6965408
KEYWORDS
EST.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 484)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
COMMENT
JOURNAL
MEDLINE
CONTACT: Soares, MB
9704477
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msocres@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized AV canal at 16.5 dpc library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-48,
>POLY_A#simple-repeat
Seq primer: M13 Forward
POLYA=yes.

Location/Qualifiers
1.484
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0p-afd-c-06-0-UI"
/clone.lib="UI-R-BJ0p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTR3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-BJ0p
library is a subcloned library derived from the UI-R-AA1,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research

Query Match 1.1%; Score 23; DB 14; Length 481;
Best Local Similarity 100.0%; Pred. NO. 1.6e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 117 a 103 c 124 g 137 t
ORIGIN
TAG-11B-UI-R-CN1
TAG-TISSUE=lungus
TAG_SEQ=TTCCG"

6: 791-806, 1996.
TAG_LIB=OI-R-BJOP
TAG_TISSUE=AV canal at 16.5 dpc
TAG_SEQ=GAACC"
BASE COUNT 86 a 113 c 100 g 185 t
ORIGIN

Query Match 1.1%; Score 23; DB 10; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2009 CCAATAGAGAAAAA 2031
DB 23 CCAATAGAGAAAAA 1

RESULT 49
BF062672/c 484 bp mRNA linear EST 16-OCT-2000
DEFINITION 7b53902.x1 NCI_CGAP_C016 Homo sapiens cDNA clone IMAGE:3320650 3',
mRNA sequence.
ACCESSION BF062672
VERSION BF062672.1 GI:10821582
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 484)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL. Send email to:
linfo@image.llnl.gov
Seq primer: -400p from Gibco
High quality sequence stop: 377.
Location/Qualifiers

FEATURES

source

1. 484
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_C016"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_C010 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 149 a 87 c 80 g 168 t
ORIGIN

Query Match 1.1%; Score 23; DB 12; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2010 CCAATAGAGAAAAA 2032
DB 26 CCAATAGAGAAAAA 4

RESULT 50
AA809972/c 497 bp mRNA linear EST 19-FEB-1998
LOCUS ca95f12.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1320047 3',
DEFINITION similar to gb:U13315 LACTOYLGLUTATHIONE LYASE (HUMAN); mRNA
sequence.

ACCESSION AA809972
VERSION AA809972.1 GI:2879378
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 497)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 2392 Std Error: 0.00
Seq primer: -40m13 fwd. Er from Amersham
High quality sequence stop: 370.
Location/Qualifiers

FEATURES

source

1. 497
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAGTGGAGCGCCGCTCTTTTCTTTTCTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73D vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 154 a 93 c 84 g 166 t
ORIGIN

Query Match 1.1%; Score 23; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2010 CCAATAGAGAAAAA 2032
DB 23 CCAATAGAGAAAAA 1

Search completed: January 11, 2003, 01:21:55
Job time : 3172 secs

|||||
Db 841 TCTATATATGACGTCGCTGTGATGGGGGAGCATGACGACAACTCAGAGAGAGG 900
QY 901 CCAACCTTACTATGTATGAGGTCATCTGCGCAAGCCAGCTGAGATCTACAGACCAT 960
Db 901 CCAACCTTACTATGTATGAGGTCATCTGCGCAAGCCAGCTGAGATCTACAGACCAT 960
QY 961 CCAATCTTCCCAAGGCGCTGAGAGACCTACCTGCTTGTGCTGATGAGAGACCTGGC 1020
Db 961 CCAATCTTCCCAAGGCGCTGAGAGACCTACCTGCTTGTGCTGATGAGAGACCTGGC 1020
QY 1021 TCGAGCCCTGTGGCCAGACTTCCGCAATGTATGATTCGTGGATTTGGAATTTCTGGC 1080
Db 1021 TCGAGCCCTGTGGCCAGACTTCCGCAATGTATGATTCGTGGATTTGGAATTTCTGGC 1080
QY 1081 CCAATTTTCAAGCTGGGTGATTAACATCACCCAGGCAAGGCGATGGGTGACACCTTT 1140
Db 1081 CCAATTTTCAAGCTGGGTGATTAACATCACCCAGGCAAGGCGATGGGTGACACCTTT 1140
QY 1141 CCACACAATGCCAGGAGTCCCTTAATGTCTCCAGGCTTGGCGCGCTTTGGCCTTA 1200
Db 1141 CCACACAATGCCAGGAGTCCCTTAATGTCTCCAGGCTTGGCGCGCTTTGGCCTTA 1200
QY 1201 TGAAGAAGTTTCTGACCTTGAGAAAGCCTGTGCGATGCCATGAATTTGCTGGCTACCG 1260
Db 1201 TGAAGAAGTTTCTGACCTTGAGAAAGCCTGTGCGATGCCATGAATTTGCTGGCTACCG 1260
QY 1261 CCAAGTCAGATTCGAGAAACAAGAGAAACCTGTGCTGATCTCTGCTACCTGGAC 1320
Db 1261 CCAAGTCAGATTCGAGAAACAAGAGAAACCTGTGCTGATCTCTGCTACCTGGAC 1320
QY 1321 TGTCCCTGAGCAATCCACTAAGAGGGTTGAGAGGCTTTGCTGCCACCTGATGACGC 1380
Db 1321 TGTCCCTGAGCAATCCACTAAGAGGGTTGAGAGGCTTTGCTGCCACCTGATGACGC 1380
QY 1381 TCAATCTCTTCTGATGATCTTCTGAGCCTTGCCTACATCTCTGAGCCTTACCTAC 1440
Db 1381 TCAATCTCTTCTGATGATCTTCTGAGCCTTGCCTACATCTCTGAGCCTTACCTAC 1440
QY 1441 TCTGTGGGTAATCACTAGATGTGAGTTGTCTCCACAGCTGCTCAAGCAGAGAGACTTT 1500
Db 1441 TCTGTGGGTAATCACTAGATGTGAGTTGTCTCCACAGCTGCTCAAGCAGAGAGACTTT 1500
QY 1501 GTGTCCATGCTTGTGTGTATAAAAACAGCTGGGAACTTATGTAGCAGACATCTCCAC 1560
Db 1501 GTGTCCATGCTTGTGTGTATAAAAACAGCTGGGAACTTATGTAGCAGACATCTCCAC 1560
QY 1561 CAGTGAACAGGGATATGCTCTTCTTCTTGTGATCTCTGCTGGGCACTTCAAG 1620
Db 1561 CAGTGAACAGGGATATGCTCTTCTTCTTGTGATCTCTGCTGGGCACTTCAAG 1620
QY 1621 AGACTTTGTGGCTGAGGCTTATTAAGCAGACAGATATCAGTGAATTTGATCCATAA 1680
Db 1621 AGACTTTGTGGCTGAGGCTTATTAAGCAGACAGATATCAGTGAATTTGATCCATAA 1680
QY 1681 ACCCTCCCTGACATGTCGCCAATGGGGAATGSAATCTTACCAAAAGGCTCACAG 1740
Db 1681 ACCCTCCCTGACATGTCGCCAATGGGGAATGSAATCTTACCAAAAGGCTCACAG 1740
QY 1741 ATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGATTCCAGTGAATTCAGAGAA 1800
Db 1741 ATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGATTCCAGTGAATTCAGAGAA 1800
QY 1801 GTGGGAACAAGGTTGATGCTTACTATGAGCTTGACCATACAGCTATGCTAATCAAG 1860
Db 1801 GTGGGAACAAGGTTGATGCTTACTATGAGCTTGACCATACAGCTATGCTAATCAAG 1860
QY 1861 AATGTGAACAAAATCTCTGACAAAAGAGCAGCTTATAGTTACAGAGGCTCCGAGC 1920
Db 1861 AATGTGAACAAAATCTCTCTGACAAAAGAGCAGCTTATAGTTACAGAGGCTCCGAGC 1920
QY 1921 TGCATTTGAATATCACTTCCCTCTGCAATTTCCATACATGAAGACTTGAACCTGTG 1980
Db 1921 TGCATTTGAATATCACTTCCCTCTGCAATTTCCATACATGAAGACTTGAACCTGTG 1980

Db 1921 TGCATTTGAATATCACTTCCCTCTGCAATTTCCATACATGAAGACTTGAACCTGTG 1980
QY 1981 AAGTCGCATCTGTTTAATACATAAATTTCCCAATATAGAAAAA 2032
Db 1981 AAGTCGCATCTGTTTAATACATAAATTTCCCAATATAGAAAAA 2032

RESULT 2

US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190, 911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 100.0%; Score 2032; DB 4; Length 2032;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGGAGGCGGAGATGCTCCAGCTGAGGAGAAATGCTTCGATTTGCTTCGCCAG 60
Db 1 GGTGGAGGCGGAGATGCTCCAGCTGAGGAGAAATGCTTCGATTTGCTTCGCCAG 60
QY 61 CCCACCTAAGCAGTCTGCCACCCCTTGAATCTGAGAGAGTGAAGCTTACTTTCA 120
Db 61 CCCACCTAAGCAGTCTGCCACCCCTTGAATCTGAGAGAGTGAAGCTTACTTTCA 120
QY 121 CAGCTTCCGAGGAGAGTCTTCTCAAGCCGCTGCAAGGCTTCCACTTCAGAC 180
Db 121 CAGCTTCCGAGGAGAGTCTTCTCAAGCCGCTGCAAGGCTTCCACTTCAGAC 180
QY 181 AATGCTACTGCTAAATAAATGAAGCTCTGCTGTTTCTGATTTCCAGATGGCATCT 240
Db 181 AATGCTACTGCTAAATAAATGAAGCTCTGCTGTTTCTGATTTCCAGATGGCATCT 240
QY 241 GCTCTATTCTTCCAGATGTACAGCCACAATAGCTCCCTGTCTATGAGGAGACACC 300
Db 241 GCTCTATTCTTCCAGATGTACAGCCACAATAGCTCCCTGTCTATGAGGAGACACC 300
QY 301 CGAGCGCATGACGCTGCTGTTCTGTTCCGCGCTCTGAGCTTCTTTTGTGGGCA 360
Db 301 CGAGCGCATGACGCTGCTGTTCTGTTCCGCGCTCTGAGCTTCTTTTGTGGGCA 360
QY 361 GCTTTTGGGAGACACCCAGATGTTTCTTACTGATGAGCCCGCTGGACGCTGTGAT 420
Db 361 GCTTTTGGGAGACACCCAGATGTTTCTTACTGATGAGCCCGCTGGACGCTGTGAT 420
QY 421 GACCTTCAAGCAGACCCGCTGGATGCTGACATGCTGTGGGAGATGATACGGGC 480
Db 421 GACCTTCAAGCAGACCCGCTGGATGCTGACATGCTGTGGGAGATGATACGGGC 480
QY 481 CGTCTTCTGTCGACATGAGGCTTTTGANCCATAGAACTGCTCCCGAGACA 540
Db 481 CGTCTTCTGTCGACATGAGGCTTTTGANCCATAGAACTGCTCCCGAGACA 540
QY 541 GTCCAGCCCTCTTTCAGTGGGAGAAAGCCGGGCTCTGTCTGTGACCTGCTGTGAT 600
Db 541 GTCCAGCCCTCTTTCAGTGGGAGAAAGCCGGGCTCTGTCTGTGACCTGCTGTGAT 600


```
QY 601 CATCCCAAGATGAAATCATATCCCGGGGCTCAGTGCAGAGCTCTCTGTGCAATCAAGACC 660
Db 601 CATCCCAAGATGAAATCATATCCCGGGGCTCAGTGCAGAGCTCTCTGTGCAATCAAGACC 660
QY 661 CTTGAGGTGTGAGAGAGGCTCGCGCTCTACAGCCAGCTGTGTGCTCAAGAGAGTGGC 720
Db 661 CTTGAGGTGTGAGAGAGGCTCGCGCTCTACAGCCAGCTGTGTGCTCAAGAGAGTGGC 720
QY 721 CTTTCAACCTGCAAGTCCCTCTTACCCGCTGCTGAAGAGCCCTCCCTCAACTGCAATAT 780
Db 721 CTTTCAACCTGCAAGTCCCTCTTACCCGCTGCTGAAGAGCCCTCCCTCAACTGCAATAT 780
QY 781 CGTGACCTGTGGCGGGAGCCCGGGGGCTGTGCGTTCGCGATGAGAGGCAAGAGGAGA 840
Db 781 CGTGACCTGTGGCGGGAGCCCGGGGGCTGTGCGTTCGCGATGAGAGGCAAGAGGAGA 840
QY 841 TGTGATGATGACAGTGCATTTGTATGAGGAGCAGATGAGCAGAACTCAAGAGAGAGA 900
Db 841 TGTGATGATGACAGTGCATTTGTATGAGGAGCAGATGAGCAGAACTCAAGAGAGAGA 900
QY 901 CCAACCCCTACATGTGATGAGCAGGTGATGCTGCCAAGCCAGCTGAGATCTAAGACCAT 960
Db 901 CCAACCCCTACATGTGATGAGCAGGTGATGCTGCCAAGCCAGCTGAGATCTAAGACCAT 960
QY 961 CCAGTCTTGGCCCAAGGCGCTGCAAGAGACGCTACCTGCTGCGCTATGAGAGCTGGC 1020
Db 961 CCAGTCTTGGCCCAAGGCGCTGCAAGAGACGCTACCTGCTGCGCTATGAGAGCTGGC 1020
QY 1021 TCGAGCCCTGTGGCCCAAGGCTTCCCGAATGATGATTTGCTGGAGATGAAATCTTGGC 1080
Db 1021 TCGAGCCCTGTGGCCCAAGGCTTCCCGAATGATGATTTGCTGGAGATGAAATCTTGGC 1080
QY 1081 CCATCTTGAAGCTGGGTGATGATACATACCCAGGAGGAGGAGGATGGTGCACAGGCTT 1140
Db 1081 CCATCTTGAAGCTGGGTGATGATACATACCCAGGAGGAGGAGGATGGTGCACAGGCTT 1140
QY 1141 CCACCAAAATGCCAGGAGTGGCCCTTAATGTCTCCAGGCTTGGGCTGGTCTTGGCCTA 1200
Db 1141 CCACCAAAATGCCAGGAGTGGCCCTTAATGTCTCCAGGCTTGGGCTGGTCTTGGCCTA 1200
QY 1201 TGAAGAGTGTCTGACTTCAGAAAGGCTGTGGCAGTGCATGATTTGGGCTGACCG 1260
Db 1201 TGAAGAGTGTCTGACTTCAGAAAGGCTGTGGCAGTGCATGATTTGGGCTGACCG 1260
QY 1261 CCACGTCAGATCTGAACAGAGACAGAGAAACCTGTTGCTGATCTTCTGTACTGAGAC 1320
Db 1261 CCACGTCAGATCTGAACAGAGACAGAGAAACCTGTTGCTGATCTTCTGTACTGAGAC 1320
QY 1321 TGTCCCTAGCAAAATCCACTAAGAGGCTTGAAGAGCTTCTCTCCACCTGGTGTGAGCC 1380
Db 1321 TGTCCCTAGCAAAATCCACTAAGAGGCTTGAAGAGCTTCTCTCCACCTGGTGTGAGCC 1380
QY 1381 TCAAGCACTTCTCTGATGCTTGAAGCTTGGAGCTTGCCTACATCTGAGGCTTAACATAG 1440
Db 1381 TCAAGCACTTCTCTGATGCTTGAAGCTTGGAGCTTGCCTACATCTGAGGCTTAACATAG 1440
QY 1441 TCTGTGGGTATCACAGTGTGATGCTTCCACACAGCTGCTCAAGCAGAGAGACTTTT 1500
Db 1441 TCTGTGGGTATCACAGTGTGATGCTTCCACACAGCTGCTCAAGCAGAGAGACTTTT 1500
QY 1501 GTGTGAGTGTGTGTGAGAAACAGAGCTGGGAGCTTGTGAGAGAGAGACATCCAC 1560
Db 1501 GTGTGAGTGTGTGTGAGAAACAGAGCTGGGAGCTTGTGAGAGAGAGACATCCAC 1560
QY 1561 CAGTGAACAGAGGATATGCTCTTCTTCTTCTGATCTTCTGAGAGAGAGACTTCAAG 1620
Db 1561 CAGTGAACAGAGGATATGCTCTTCTTCTTCTGATCTTCTGAGAGAGAGACTTCAAG 1620
QY 1621 AGACTTGTGGCTGGAGGCTATTAAGCAGACAGTATCAAGTGAATGATGATCAATA 1680
Db 1621 AGACTTGTGGCTGGAGGCTATTAAGCAGACAGTATCAAGTGAATGATGATCAATA 1680
```

```
QY 1681 ACCTCCCTGTCCACATCTTCCCATGGSAGATGATCTTTTCACCAAGAGAGCTCACCAG 1740
Db 1681 ACCTCCCTGTCCACATCTTCCCATGGSAGATGATCTTTTCACCAAGAGAGCTCACCAG 1740
QY 1741 ATTTTCACAGAGATGCAAAATTCAGAGCCCTTGAGTTCCTCAATGATTAAGAGAGAA 1800
Db 1741 ATTTTCACAGAGATGCAAAATTCAGAGCCCTTGAGTTCCTCAATGATTAAGAGAGAA 1800
QY 1801 GTGGGAACAAGGTGTGATGCTTCTTATGAGCTTGACCATCAGCTATTCGTAATCAGA 1860
Db 1801 GTGGGAACAAGGTGTGATGCTTCTTATGAGCTTGACCATCAGCTATTCGTAATCAGA 1860
QY 1861 AATATGAACAAATCTCTGACAAAGAGCAGCTTATGATGATGAGAGGCTGCTGGG 1920
Db 1861 AATATGAACAAATCTCTGACAAAGAGCAGCTTATGATGATGAGAGGCTGCTGGG 1920
QY 1921 TGCATTTGAATATCACTTCCCTCTGCAATTTCCATCAGATGAGAGACTTTGACCTGTG 1980
Db 1921 TGCATTTGAATATCACTTCCCTCTGCAATTTCCATCAGATGAGAGACTTTGACCTGTG 1980
QY 1981 AAGCTGCATCTGTATTAATAATCCCAATTAAGAAAAA 2032
Db 1981 AAGCTGCATCTGTATTAATAATAATCCCAATTAAGAAAAA 2032

RESULT 3
US-09-045-284A-7
; Sequence 7, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-045-284A-7

Query Match 1.8%; Score 37; DB 4; Length 37;
Best Local Similarity 100.0%; Pred No. 4.3e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 884 AACTCAAGAAGAGGACCAACCTACTATGATGTC 920
Db 1 AACTCAAGAAGAGGACCAACCTACTATGATGTC 37

RESULT 4
US-09-190-911-7
; Sequence 7, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 37
```

TYPE: DNA
ORGANISM: H. sapiens
US-09-190-911-7

Query Match
Best Local Similarity 100.0%; Pred. No. 4.3e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 884 AACTCAAGAGAGAGACCACTACTATGTGATGC 920
DB 1 AACTCAAGAGAGAGACCACTACTATGTGATGC 37

US-07-867-106-4/c
Sequence 4, Application US/07867106
Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Feehey, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-9100
TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

LENGTH: 3138 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single

TOPOLOGY: Circular
MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO
US-07-867-106-4

Query Match
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATATAGAAAAA 2032
DB 2030 AATATAGAAAAA 2009

RESULT 6
5340934-5
Patent No. 5340934

APPLICANT: TERMININE, JOHN D., YOUNG, MARIAN F., FISHER, LARRY W.

ROBEY, PAMELA G.
TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/432,044
FILING DATE: 03-NOV-1989
SEQ ID NO: 5
LENGTH: 1493

Query Match
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAAAAA 2032
DB 1440 AATAGAAAAA 1460

US-08-171-385-16/c
Sequence 16, Application US/08171385
Patent No. 5527884

GENERAL INFORMATION:

APPLICANT: Mary E. Russell
APPLICANT: Ulrike Utans
TITLE OF INVENTION: Mediators of Chronic Allotransplant
TITLE OF INVENTION: Rejection
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/171,385
FILING DATE:

CLASSIFICATION: 514
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/006001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:

LENGTH: 295
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
US-08-171-385-16

Query Match
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 AATAGAAAAA 2032
DB 27 AATAGAAAAA 8

```

RESULT 8
US-08-361-441B-16/c
; Sequence 16, Application US/08361441B
; Patent No. 6077948
; GENERAL INFORMATION:
; APPLICANT: Russell, Mary E.
; APPLICANT: Utans, Ulrike
; TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,441B
; FILING DATE: 21-DEC-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/171,385
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/014001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 295 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-361-441B-16
Query Match 1.0%; Score 20; DB 3; Length 295;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 27 ATAGAGAAAAA 8

```

```

RESULT 9
US-08-171-385-12/c
; Sequence 12, Application US/08171385
; Patent No. 5527884
; GENERAL INFORMATION:
; APPLICANT: Mary E. Russell
; APPLICANT: Ulrike Utans
; TITLE OF INVENTION: Mediators of Chronic Allograft
; TITLE OF INVENTION: Rejection
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)

```

```

; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,385
; FILING DATE:
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/006001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-171-385-12
Query Match 1.0%; Score 20; DB 1; Length 362;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 35 ATAGAGAAAAA 16

```

```

RESULT 10
US-08-361-441B-12/c
; Sequence 12, Application US/08361441B
; Patent No. 6077948
; GENERAL INFORMATION:
; APPLICANT: Russell, Mary E.
; APPLICANT: Utans, Ulrike
; TITLE OF INVENTION: MEDIATORS OF CHRONIC ALLOGRAFT REJECTION
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,441B
; FILING DATE: 21-DEC-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/171,385
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/014001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

; TOPOLOGY: linear
; US-08-361-441B-12
;
; Query Match
; Best Local Similarity 100.0%; Score 20; DB 3; Length 362;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2013 ATAGAGAAAAA 2032
; 35 ATAGAGAAAAA 16
;
; RESULT 11
; US-09-247-155-152
; Sequence 152, Application US/09247155A
; Patent No. 6312822
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bouquellet, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET 021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 152
; LENGTH: 367
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 70..321
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 70..234
; OTHER INFORMATION: Von Heljne matrix
; OTHER INFORMATION: score 4.03999990463257
; OTHER INFORMATION: seq AVCAALLASHPTA/EV
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 364..369
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 375..387
; US-09-247-155-152
;
; Query Match
; Best Local Similarity 100.0%; Score 20; DB 4; Length 387;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2011 AATAGAGAAAAA 2030
; 368 AATAGAGAAAAA 387
;
; RESULT 12
; US-09-336-536-26
; Sequence 26, Application US/09336536
; Patent No. 6406884
; GENERAL INFORMATION:
; APPLICANT: Leidy, K.
; APPLICANT: McKay, C.
; APPLICANT: Bossone, S.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
```

```

; FILE REFERENCE: 7853-144
; CURRENT APPLICATION NUMBER: US/09/336,536
; CURRENT FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1869
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: all "n" positions
; OTHER INFORMATION: n=a, c, g, or t
; US-09-336-536-26
;
; Query Match
; Best Local Similarity 100.0%; Score 20; DB 4; Length 1869;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2013 ATAGAGAAAAA 2032
; 1840 ATAGAGAAAAA 1859
;
; RESULT 13
; US-09-360-545-68
; Sequence 68, Application US/09360545
; Patent No. 6429014
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Bohmann, Jorg
; APPLICANT: Steele, Christopher L
; APPLICANT: Phillips, Michael A
; TITLE OF INVENTION: MONOTERPENE SYNTHASES FROM GRAND FIR (ABIES GRANDIS)
; FILE REFERENCE: wslu13885
; CURRENT APPLICATION NUMBER: US/09/360,545
; CURRENT FILING DATE: 1999-07-26
; EARLIER APPLICATION NUMBER: 60/052,249
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: PCT/US98/14528
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 68
; LENGTH: 2429
; TYPE: DNA
; ORGANISM: Abies grandis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (35)..(1945)
; US-09-360-545-68
;
; Query Match
; Best Local Similarity 100.0%; Score 20; DB 4; Length 2429;
; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 2013 ATAGAGAAAAA 2032
; 2407 ATAGAGAAAAA 2426
;
; RESULT 14
; US-09-813-133A-3
; Sequence 3, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: C1001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
```

Mon Jan 13 09:13:56 2003

us-09-816-825-1.oli20.rni

Page 7

; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-3

Query Match 1.0%; Score 20; DB 4; Length 55827;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2011 AATATAGAAAAAAAAAAAAA 2030
|||||
Db 24049 AATATAGAAAAAAAAAAAAA 24068

Search completed: January 10, 2003, 21:40:17
Job time : 205 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 19:13:52 ; Search time 95 Seconds

(without alignments)
9413.579 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 2032
Sequence: 1 ggctcagagccagatgctct.....ataagaaaaaaaaaaaaa 2032

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 389086 seqs, 220051671 residues

Word size : 20

Total number of hits satisfying chosen parameters: 47

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published.Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 2032 | 100.0 | 2032 | 9 | US-10-007-262-2 |
| 2 | 2032 | 100.0 | 2032 | 10 | US-09-816-825-1 |
| 3 | 505 | 24.9 | 505 | 10 | US-09-998-598-2595 |
| 4 | 505 | 24.9 | 517 | 10 | US-09-919-580-194 |
| 5 | 286 | 14.1 | 389 | 10 | US-09-919-580-264 |
| 6 | 59 | 2.9 | 2544 | 10 | US-09-927-602-1 |
| 7 | 59 | 2.9 | 48436 | 10 | US-09-927-602-38 |
| 8 | 37 | 1.8 | 37 | 9 | US-10-007-262-7 |
| 9 | 37 | 1.8 | 37 | 10 | US-09-816-825-7 |
| 10 | 22 | 1.1 | 1782 | 10 | US-09-812-800A-1462 |
| 11 | 22 | 1.1 | 2557 | 10 | US-09-812-102-10 |
| 12 | 21 | 1.0 | 242 | 10 | US-09-960-352-3651 |
| 13 | 21 | 1.0 | 260 | 10 | US-09-983-965-5796 |
| 14 | 21 | 1.0 | 280 | 10 | US-09-960-352-4352 |
| 15 | 21 | 1.0 | 411 | 9 | US-09-954-531-265 |
| 16 | 21 | 1.0 | 411 | 10 | US-09-880-107-1438 |
| 17 | 21 | 1.0 | 785 | 10 | US-09-770-445-859 |
| 18 | 21 | 1.0 | 986 | 10 | US-09-770-445-263 |
| 19 | 21 | 1.0 | 998 | 10 | US-09-925-301-297 |

| | | | | | | |
|----|----|-----|------|----|---------------------|--------------------|
| 20 | 21 | 1.0 | 1263 | 9 | US-10-016-157A-89 | Sequence 89, Appl |
| 21 | 21 | 1.0 | 1540 | 10 | US-09-998-598-352 | Sequence 352, App |
| 22 | 21 | 1.0 | 1469 | 10 | US-09-919-770-1 | Sequence 1, Appl1 |
| 23 | 21 | 1.0 | 1792 | 10 | US-09-925-301-259 | Sequence 259, App |
| 24 | 21 | 1.0 | 2000 | 9 | US-09-938-842A-4722 | Sequence 4722, Ap |
| 25 | 21 | 1.0 | 2000 | 9 | US-09-925-301-572 | Sequence 572, Ap |
| 26 | 20 | 1.0 | 101 | 10 | US-09-983-965-3358 | Sequence 3358, App |
| 27 | 20 | 1.0 | 161 | 10 | US-09-867-701-9309 | Sequence 9309, Ap |
| 28 | 20 | 1.0 | 176 | 10 | US-09-983-965-2866 | Sequence 2866, Ap |
| 29 | 20 | 1.0 | 326 | 10 | US-09-960-352-5131 | Sequence 8131, Ap |
| 30 | 20 | 1.0 | 394 | 10 | US-09-960-352-5062 | Sequence 5062, Ap |
| 31 | 20 | 1.0 | 398 | 10 | US-09-960-352-11099 | Sequence 11099, A |
| 32 | 20 | 1.0 | 403 | 10 | US-09-960-352-7945 | Sequence 7945, Ap |
| 33 | 20 | 1.0 | 410 | 10 | US-09-919-580-294 | Sequence 294, App |
| 34 | 20 | 1.0 | 430 | 10 | US-09-960-352-9747 | Sequence 9747, App |
| 35 | 20 | 1.0 | 474 | 10 | US-09-834-975-616 | Sequence 616, App |
| 36 | 20 | 1.0 | 477 | 9 | US-09-938-842A-3890 | Sequence 3890, Ap |
| 37 | 20 | 1.0 | 498 | 10 | US-09-764-846-16 | Sequence 16, Appl |
| 38 | 20 | 1.0 | 517 | 10 | US-09-917-800A-306 | Sequence 306, App |
| 39 | 20 | 1.0 | 554 | 10 | US-09-919-580-462 | Sequence 462, App |
| 40 | 20 | 1.0 | 615 | 10 | US-09-867-550-1485 | Sequence 1485, App |
| 41 | 20 | 1.0 | 630 | 10 | US-09-925-300-548 | Sequence 548, App |
| 42 | 20 | 1.0 | 1117 | 10 | US-09-925-302-240 | Sequence 240, App |
| 43 | 20 | 1.0 | 1389 | 10 | US-09-820-893-14 | Sequence 14, Appl |
| 44 | 20 | 1.0 | 1965 | 9 | US-09-938-842A-4048 | Sequence 4048, Ap |
| 45 | 20 | 1.0 | 2429 | 9 | US-10-041-007-27 | Sequence 27, Appl |
| 46 | 20 | 1.0 | 2441 | 10 | US-09-822-830A-422 | Sequence 422, App |
| 47 | 20 | 1.0 | 3492 | 10 | US-09-529-063-78 | Sequence 78, Appl |

ALIGNMENTS

RESULT 1
US-10-007-262-2
; Sequence 2, Application US/10007262
; Patent No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007/262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-2
Query Match 100.0%; Score 2032; DB 9; Length 2032;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCTCAGGCGCAGAGATGCCCTCCAGCTCTGGGGGAAATGCTTCTCAATTGCTTCTCCAG 60
1 GGCTCAGGCGCAGAGATGCCCTCCAGCTCTGGGGGAAATGCTTCTCAATTGCTTCTCCAG 60
DB 1 GGCTCAGGCGCAGAGATGCCCTCCAGCTCTGGGGGAAATGCTTCTCAATTGCTTCTCCAG 60
QY 61 CCACCTTCAGCAGCTGTCCACCCCTTGTAGTCTGCAGCTGTAAAGCTGTACTTTCA 120
61 CCACCTTCAGCAGCTGTCCACCCCTTGTAGTCTGCAGCTGTAAAGCTGTACTTTCA 120
DB 61 CCACCTTCAGCAGCTGTCCACCCCTTGTAGTCTGCAGCTGTAAAGCTGTACTTTCA 120
QY 121 CAGCTTCCTGGGAGCGAGTCTTTCAGCCCGTGTTCGACAGTCTTCACCTTCAGCAC 180
121 CAGCTTCCTGGGAGCGAGTCTTTCAGCCCGTGTTCGACAGTCTTCACCTTCAGCAC 180
DB 121 CAGCTTCCTGGGAGCGAGTCTTTCAGCCCGTGTTCGACAGTCTTCACCTTCAGCAC 180
QY 181 AATGTAAGTCTCTAAATAATGAAGTCTGCTGTTTGGTTTCCAGATGGCCACTT 240
181 AATGTAAGTCTCTAAATAATGAAGTCTGCTGTTTGGTTTCCAGATGGCCACTT 240

181 AATCTCTGCTCTTAAAAAATGAGCTCCGCTGTTCTGTTCCAGATGAGCACTTT 240
241 GGCCTCTATCTTCCACATGTCACGACCAACATCAGCTCCCTGTCATGAGGACAGCC 300
241 GGCCTCTATCTTCCACATGTCACGACCAACATCAGCTCCCTGTCATGAGGACAGCC 300
301 CGAGCGCATGACGATGCTGTTCTGTTCTGCTGCGCTGCTGCTCTTTTGTGGGCA 360
301 CGAGCGCATGACGATGCTGTTCTGTTCTGCTGCGCTGCTGCTCTTTTGTGGGCA 360
361 GCTTTTGGGAGACCCAGATGTTTCTACCTGATGAGACCCGCGCTGGACGCTGAT 420
361 GCTTTTGGGAGACCCAGATGTTTCTACCTGATGAGACCCGCGCTGGACGCTGAT 420
421 GACCTTGAAGCAGACAGCCGCTGATGCTGACATGCTGCTGCTGCTGCTGATCAGCC 480
421 GACCTTGAAGCAGACAGCCGCTGATGCTGACATGCTGCTGCTGCTGATCAGCC 480
481 CGTCTCTGTCGACATGAGCGCTTGTGATGCTGATGAGACCTGCTGCTGCTGAGCA 540
481 CGTCTCTGTCGACATGAGCGCTTGTGATGCTGATGAGACCTGCTGCTGCTGAGCA 540
541 GTCCAGCCTCTTCTAGTGGAGAAACAGCCGCGCTGCTGCTGCTGCTGCTGCTGAT 600
541 GTCCAGCCTCTTCTAGTGGAGAAACAGCCGCGCTGCTGCTGCTGCTGCTGCTGAT 600
601 CATCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
601 CATCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
661 CTTGAGTGTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
661 CTTGAGTGTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
721 CTTCTTCAACCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 780
721 CTTCTTCAACCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 780
781 CGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
781 CGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
841 TCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
841 TCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
901 CCAACCTCTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
901 CCAACCTCTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
961 CCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
961 CCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
1021 TCGAGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
1021 TCGAGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
1081 CCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1081 CCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
1141 CCAACCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
1141 CCAACCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
1201 TGAAGAAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
1201 TGAAGAAGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
1261 CCAAGTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
1261 CCAAGTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320

1261 CCAAGTCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
1321 TGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAGGCTTGTGCTGCTGCTGCTGCTGCTG 1380
1321 TGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAGGCTTGTGCTGCTGCTGCTGCTGCTG 1380
1381 TCACTGATCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
1381 TCACTGATCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
1441 TCTGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
1441 TCTGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
1501 GTGTCCATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1501 GTGTCCATGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1561 CAGTGAACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
1561 CAGTGAACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
1621 AGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
1621 AGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
1681 ACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
1681 ACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
1741 ATTTTCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
1741 ATTTTCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
1801 GTGGGAACAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
1801 GTGGGAACAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
1861 AATATGAACAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
1861 AATATGAACAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
1921 TGCATTTGAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
1921 TGCATTTGAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
1981 AAGCTGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2032
1981 AAGCTGCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2032

RESULT 2
US-09-816-825-1
; Sequence 1, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816, 825
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045, 284
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 1
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-825-1

Query Match 100.0%; Score 2032; DB 10; Length 2032;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCTCGAGGCGAGGATGCTCCATGCTGGGGGAAATGCTTCCATTTCTCTCCAG 60
DB 1 GGCTCGAGGCGAGGATGCTCCATGCTGGGGGAAATGCTTCCATTTCTCTCCAG 60

QY 61 CCAAGCTCAGAGGATGCTCCAGGCTTGAAGTCTCAGAGCTTTAAAGTGTACTTTCA 120
DB 61 CCAAGCTCAGAGGATGCTCCAGGCTTGAAGTCTCAGAGCTTTAAAGTGTACTTTCA 120

QY 121 CAGGCTCTGGAGGAGGATGCTTCTCAAGGCGGCTTGAAGGCTTCCAGTTCAGAC 180
DB 121 CAGGCTCTGGAGGAGGATGCTTCTCAAGGCGGCTTGAAGGCTTCCAGTTCAGAC 180

QY 181 AATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCCATCTT 240
DB 181 AATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCCATCTT 240

QY 241 GGCTCTATCTTCCACATGTACAGCCACACATAGCTCCCTGTCTATGAAGGACAGCC 300
DB 241 GGCTCTATCTTCCACATGTACAGCCACACATAGCTCCCTGTCTATGAAGGACAGCC 300

QY 301 CGAGCGCATGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 301 CGAGCGCATGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360

QY 361 GCTTTTGGGCGACCAACAGATGTTTCTACCTGATGAGGCCCTGCGACAGTGGAT 420
DB 361 GCTTTTGGGCGACCAACAGATGTTTCTACCTGATGAGGCCCTGCGACAGTGGAT 420

QY 421 GACCTTCAAGCAGAGCAAGGCTGATGCTGACATGAGCTGCTGAGGATCTGATAGAGG 480
DB 421 GACCTTCAAGCAGAGCAAGGCTGATGCTGACATGAGCTGCTGAGGATCTGATAGAGG 480

QY 481 CGCTTCTTGGCGACATGAGCGTCTTGTGATGCTTACATGAGAACCTGGTCCCGAGACA 540
DB 481 CGCTTCTTGGCGACATGAGCGTCTTGTGATGCTTACATGAGAACCTGGTCCCGAGACA 540

QY 541 GTCAGGCTCTTTCAGTGGGAGAGACAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 541 GTCAGGCTCTTTCAGTGGGAGAGACAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTG 600

QY 601 CATCCACAGAGTGAATCATCCCGGCGCTCACTGCAAGGCTCTGTGAGTCAAGAGCC 660
DB 601 CATCCACAGAGTGAATCATCCCGGCGCTCACTGCAAGGCTCTGTGAGTCAAGAGCC 660

QY 661 CTTTGAAGTGGTGAAGAGGCTGCTGCTTCTTACAGCCAGTGGTGGTCAAGAGTGGC 720
DB 661 CTTTGAAGTGGTGAAGAGGCTGCTGCTTCTTACAGCCAGTGGTGGTCAAGAGTGGC 720

QY 721 CTTTTCACCTGCGAGTCCCTCTACCGCTGCTGAAGAGCCCTCCCTCAACCTGCATAT 780
DB 721 CTTTTCACCTGCGAGTCCCTCTACCGCTGCTGAAGAGCCCTCCCTCAACCTGCATAT 780

QY 781 CGTGACCTGGTCCGGAGCCCGGCGCGTGTCCGTTCCCGAGAGCGACAAAGGAGGA 840
DB 781 CGTGACCTGGTCCGGAGCCCGGCGCGTGTCCGTTCCCGAGAGCGACAAAGGAGGA 840

QY 841 TCTCATGATGACAGTGCATGTTGATGAGGAGCAGATGAGCAGAACTCAAGAGAGGA 900
DB 841 TCTCATGATGACAGTGCATGTTGATGAGGAGCAGATGAGCAGAACTCAAGAGAGGA 900

QY 901 CCAAGCTACTATGTATGATGAGGCTATCTGCCAAAGCCAGCTGAGATCTTCAAGAGCAT 960
DB 901 CCAAGCTACTATGTATGATGAGGCTATCTGCCAAAGCCAGCTGAGATCTTCAAGAGCAT 960

QY 961 CCAAGCTTGGCCCAAGGCGCTGCAAGAGGCTTACCTGCTGTGTGAGGAGCTGGCC 1020
DB 961 CCAAGCTTGGCCCAAGGCGCTGCAAGAGGCTTACCTGCTGTGTGAGGAGCTGGCC 1020

QY 1021 TCGAGCCCTGTGGCCAGACTTCCGAAATGTATGAATTCGTGGATTTGAATCTTGCC 1080

DB 1021 TCGAGCCCTGTGGCCAGACTTCCGAAATGTATGAATTCGTGGATTTGAATCTTGCC 1080

QY 1081 CCAATCTTCAGAGCTGGGTGATATACATACACCCGAGGCAAGGGCATGGTACACAGCTTT 1140
DB 1081 CCAATCTTCAGAGCTGGGTGATATACATACACCCGAGGCAAGGGCATGGTACACAGCTTT 1140

QY 1141 CCAACAAATGCCAGGAGATGCCCTTAATGTCTCCAGGCTTGGCGCTGTCTTGGCCCTA 1200
DB 1141 CCAACAAATGCCAGGAGATGCCCTTAATGTCTCCAGGCTTGGCGCTGTCTTGGCCCTA 1200

QY 1201 TGAAGAGTTTCTCGACTTCAGAAAGCCTGTGGAGATGTCATGAATTTCTGGGAGCCG 1260
DB 1201 TGAAGAGTTTCTCGACTTCAGAAAGCCTGTGGAGATGTCATGAATTTCTGGGAGCCG 1260

QY 1261 CCAAGCTCAGATCTGAACAAAGAGAGAAACCTGTTGCTGATCTTCTGCTACCTGAGC 1320
DB 1261 CCAAGCTCAGATCTGAACAAAGAGAGAAACCTGTTGCTGATCTTCTGCTACCTGAGC 1320

QY 1321 TGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTTGTCTGCCACCTGGTGTACGC 1380
DB 1321 TGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTTGTCTGCCACCTGGTGTACGC 1380

QY 1381 TCACTCACTTCTCTGAATGCTCTGAGCCTTGGCTTACATCTCTGAGCCTTAACATCATG 1440
DB 1381 TCACTCACTTCTCTGAATGCTCTGAGCCTTGGCTTACATCTCTGAGCCTTAACATCATG 1440

QY 1441 TCTGTGGGATTCACACTGAGTGTGAGTGTGTCTCACAGCTGCTCAAGCAAGAGACTTTT 1500
DB 1441 TCTGTGGGATTCACACTGAGTGTGAGTGTGTCTCACAGCTGCTCAAGCAAGAGACTTTT 1500

QY 1501 GTGAGCATGCTTGTGCTGAGAAACAGACGAGGAGCCTTAATGTAGAGCAGCATCCAC 1560
DB 1501 GTGAGCATGCTTGTGCTGAGAAACAGACGAGGAGCCTTAATGTAGAGCAGCATCCAC 1560

QY 1561 CAGTGAACAGGATGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
DB 1561 CAGTGAACAGGATGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620

QY 1621 AGACTTGTGGCTGGAGGCTATTAGAGCAGACAGATCAGTGGATTTGATCCATTA 1680
DB 1621 AGACTTGTGGCTGGAGGCTATTAGAGCAGACAGATCAGTGGATTTGATCCATTA 1680

QY 1681 ACCTCCCTGTCACATCTTCCCAATGGGAAATGATCTTTCACAAAGAGCTCACACAC 1740
DB 1681 ACCTCCCTGTCACATCTTCCCAATGGGAAATGATCTTTCACAAAGAGCTCACACAC 1740

QY 1741 ATTTTCCACAGATGCAATTTGTAGCCCTTGGAGTTCCTGAGTTCCTGATTCAGAGAGGA 1800
DB 1741 ATTTTCCACAGATGCAATTTGTAGCCCTTGGAGTTCCTGAGTTCCTGATTCAGAGAGGA 1800

QY 1801 GTGGGAACAAGTTGATGCTACTTATGAGCTTGAAGCTTCAACAGCTATCGTAATCAGA 1860
DB 1801 GTGGGAACAAGTTGATGCTACTTATGAGCTTGAAGCTTCAACAGCTATCGTAATCAGA 1860

QY 1861 AATATGAACAAATCTCTGACAAAGAGCAAGCTTCTTACCTTCAAGGCTGCTGGGC 1920
DB 1861 AATATGAACAAATCTCTGACAAAGAGCAAGCTTCTTACCTTCAAGGCTGCTGGGC 1920

QY 1921 TGCATTTGAATATCAGTCCCTCTGATTTTCCATCAGATGAGAGCTTTGACCTGTG 1980
DB 1921 TGCATTTGAATATCAGTCCCTCTGATTTTCCATCAGATGAGAGCTTTGACCTGTG 1980

QY 1981 AAGCTGCATCTGTATTAATAATAATAATCCCAATTAAGAAAAA 2032
DB 1981 AAGCTGCATCTGTATTAATAATAATAATCCCAATTAAGAAAAA 2032

RESULT 3
US-09-998-598-2595
; Sequence 2595, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:

```

; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2595
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-2595

Query Match
Best Local Similarity 100.0%; Pred. No. 4e-204;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1355 GCGTTGTGCGACCTGGTGTGACCTCAGTCTGCTGTAATGCTTGTGAGCCTTGC 1414
Db 1 GGGTTTGGTGGCCACCTGGTGTGACCTCAGTCTGCTGTAATGCTTGTGAGCCTTGC 60
QY 1415 CTACATCTGTGAGCCTTAACATCATGTCTGTGGGTATCAGACTGAGTGTGCTGCC 1474
Db 61 CTACATCTGTGAGCCTTAACATCATGTCTGTGGGTATCAGACTGAGTGTGCTGCC 120
QY 1475 AACGCTGCTCAGCAGAGAGCACTTTTGTGTCATGCTTGTGTCTAGAAAACAGACTGGGG 1534
Db 121 AACGCTGCTCAGCAGAGAGCACTTTTGTGTCATGCTTGTGTCTAGAAAACAGACTGGGG 180
QY 1535 AACCTTATGTAGCAGACATCCACAGAGGAAAGGGATTTGCTTCTTCTTCTTCTT 1594
Db 181 AACCTTATGTAGCAGACATCCACAGAGGAAAGGGATTTGCTTCTTCTTCTTCTT 240
QY 1595 GATCTTCTCTCTGCGAGACTTTCAGAGACTTGTGGCTTGGAGCCTTATTAAGCAGAC 1654
Db 241 GATCTTCTCTCTGCGAGACTTTCAGAGACTTGTGGCTTGGAGCCTTATTAAGCAGAC 300
QY 1655 ACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGATG 1714
Db 301 ACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGATG 360
QY 1715 GATCTTTCACCAAGAGCTACACAGACTTTTCCACAGAGATGCAAAATTCGAGCCCTTGG 1774
Db 361 GATCTTTCACCAAGAGCTACACAGACTTTTCCACAGAGATGCAAAATTCGAGCCCTTGG 420
QY 1775 AGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGTGTGATGCTTATGAGCTT 1834
Db 421 AGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGTGTGATGCTTATGAGCTT 480
QY 1835 GACCATCAGACGATTCGCTATTCAG 1859
Db 481 GACCATCAGACGATTCGCTATTCAG 505

RESULT 4
US-09-919-580-194
; Sequence 194, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 194
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-194

Query Match
Best Local Similarity 100.0%; Pred. No. 4e-204;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1355 GCGTTGTGCGACCTGGTGTGACCTCAGTCTGCTGTAATGCTTGTGAGCCTTGC 1414
Db 13 GGGTTTGGTGGCCACCTGGTGTGACCTCAGTCTGCTGTAATGCTTGTGAGCCTTGC 72
QY 1415 CTACATCTGTGAGCCTTAACATCATGTCTGTGGGTATCAGACTGAGTGTGCTGCC 1474
Db 73 CTACATCTGTGAGCCTTAACATCATGTCTGTGGGTATCAGACTGAGTGTGCTGCC 132
QY 1475 AACGCTGCTCAGCAGAGAGCACTTTTGTGTCATGCTTGTGTCTAGAAAACAGACTGGGG 1534
Db 133 AACGCTGCTCAGCAGAGAGCACTTTTGTGTCATGCTTGTGTCTAGAAAACAGACTGGGG 192
QY 1535 AACCTTATGTAGCAGACATCCACAGAGGAAAGGGATTTGCTTCTTCTTCTTCTT 1594
Db 193 AACCTTATGTAGCAGACATCCACAGAGGAAAGGGATTTGCTTCTTCTTCTTCTT 252
QY 1595 GATCTTCTCTCTGCGAGACTTTCAGAGACTTGTGGCTTGGAGCCTTATTAAGCAGAC 1654
Db 253 GATCTTCTCTCTGCGAGACTTTCAGAGACTTGTGGCTTGGAGCCTTATTAAGCAGAC 312
QY 1655 ACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAAATG 1714
Db 313 ACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAAATG 372
QY 1715 GATCTTTCACCAAGAGCTCACCAGACTTTTCCACAGAGATGCAAAATTCGAGCCCTTGG 1774
Db 373 GATCTTTCACCAAGAGCTCACCAGACTTTTCCACAGAGATGCAAAATTCGAGCCCTTGG 432
QY 1775 AGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGTGTGATGCTTATGAGCTT 1834
Db 433 AGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGTGTGATGCTTATGAGCTT 492
QY 1835 GACCATCAGACGATTCGCTATTCAG 1859
Db 493 GACCATCAGACGATTCGCTATTCAG 517

RESULT 5
US-09-919-580-264
; Sequence 264, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-264

Query Match
Best Local Similarity 100.0%; Pred. No. 1.4e-111;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1732 CTCACACAGATTTTCACAGAGATGCAAAATTCGAGCCTTGGAGTGTCCAGTGATTC 1791
```

```

; ORGANISM: Homo Sapien
US-09-927-602-38
Query Match
Best Local Similarity 100.0%; Pred. No. 1.2e-15;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13 CTGACCGCAGCATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGTCCAGGGATTCA 72
QY 1792 AGGAGAAATGGGAAACAGATGGATGCTACTACTATAGACTTGAACACAGCTATCG 1851
Db 73 AGGAGAAATGGGAAACAGATGGATGCTACTACTATAGACTTGAACACAGCTATCG 132
QY 1852 GTAATCAGAAATATGAAACAAATCTCTGCACAAAGAGCAAGCTCTTAAGTTCACAGGG 1911
Db 133 GTAATCAGAAATATGAAACAAATCTCTGCACAAAGAGCAAGCTCTTAAGTTCACAGGG 192
QY 1912 TGCCCTGGCGTCGATTTAATATCATCTCCCTCTGCATTTTCCATCATAGAGACTT 1971
Db 193 TGCCCTGGCGTCGATTTAATATCATCTCCCTCTGCATTTTCCATCATAGAGACTT 252
QY 1972 TGACCTGTGAAGCTGCCATCTGTTAATACTAAATTCOCAATAG 2017
Db 253 TGACCTGTGAAGCTGCCATCTGTTAATACTAAATTCOCAATAG 298

RESULT 6
US-09-927-602-1
; Sequence 1, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2544
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (693)...(1877)
US-09-927-602-1
Query Match
Best Local Similarity 2.9%; Score 59; DB 10; Length 2544;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 678 AGGCTCCCGCTCCTACAGCCAGCTGTGTCTCAAGAGGCGCTTCTTAACCTGCAG 736
Db 1180 AGGCTCCCGCTCCTACAGCCAGCTGTGTCTCAAGAGGCGCTTCTTAACCTGCAG 1238

RESULT 7
US-09-927-602-38
; Sequence 38, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 48436
; TYPE: DNA
```

```

; ORGANISM: Homo Sapien
US-09-927-602-38
Query Match
Best Local Similarity 2.9%; Score 59; DB 10; Length 48436;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13 CTGACCGCAGCATTTTCCACAGAGATGCAAAATTCGAGCCCTTGAGTCCAGGGATTCA 72
QY 1792 AGGAGAAATGGGAAACAGATGGATGCTACTACTATAGACTTGAACACAGCTATCG 1851
Db 73 AGGAGAAATGGGAAACAGATGGATGCTACTACTATAGACTTGAACACAGCTATCG 132
QY 1852 GTAATCAGAAATATGAAACAAATCTCTGCACAAAGAGCAAGCTCTTAAGTTCACAGGG 1911
Db 133 GTAATCAGAAATATGAAACAAATCTCTGCACAAAGAGCAAGCTCTTAAGTTCACAGGG 192
QY 1912 TGCCCTGGCGTCGATTTAATATCATCTCCCTCTGCATTTTCCATCATAGAGACTT 1971
Db 193 TGCCCTGGCGTCGATTTAATATCATCTCCCTCTGCATTTTCCATCATAGAGACTT 252
QY 1972 TGACCTGTGAAGCTGCCATCTGTTAATACTAAATTCOCAATAG 2017
Db 253 TGACCTGTGAAGCTGCCATCTGTTAATACTAAATTCOCAATAG 298

RESULT 6
US-10-007-262-7
; Sequence 7, Application US/10007262
; Patent No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107C1P
; CURRENT APPLICATION NUMBER: US/10/007,262
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-7
Query Match
Best Local Similarity 1.8%; Score 37; DB 9; Length 37;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 884 AAACCTCAGAGAGGAGGACCAACCTACTATGTGATGC 920
Db 1 AAACCTCAGAGAGGAGGACCAACCTACTATGTGATGC 37

RESULT 9
US-09-816-825-7
; Sequence 7, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-825-7
Query Match
Best Local Similarity 1.8%; Score 37; DB 10; Length 37;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 884 AAACCTCAGAGAGGAGGACCAACCTACTATGTGATGC 920
Db 1 AAACCTCAGAGAGGAGGACCAACCTACTATGTGATGC 37
```

```
RESULT 10
; Sequence 1462, Application us/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1462
; LENGTH: 1782
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 U39208
US-09-917-800A-1462

Query Match          1.1%; Score 22; DB 10; Length 1782;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATTAAGAAAAA 2032
Db 1756 AATTAAGAAAAA 1777

RESULT 11
; Sequence 10, Application US/09812102
; Patent No. US20020055179A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E
; TITLE OF INVENTION: No. US20020055179A1 G-Protein Coupled Receptor Homologs
; FILE REFERENCE: 5800-41 035800/183478
; CURRENT APPLICATION NUMBER: US/09/812,102
; CURRENT FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/364,769
; PRIOR FILING DATE: 1999-07-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 2557
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; OTHER INFORMATION: GPCR-SECRETIN
; NAME/KEY: misc_feature
; LOCATION: (1)..(2557)
; OTHER INFORMATION: n = a, t, c, or g
US-09-812-102-10

Query Match          1.1%; Score 22; DB 10; Length 2557;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATTAAGAAAAA 2032
Db 2500 AATTAAGAAAAA 2521

RESULT 12
; Sequence 3651, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511 006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3651
; LENGTH: 242
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 16-LIB3058-047-Q1-K1-D11
US-09-960-352-3651

Query Match          1.0%; Score 21; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATTAAGAAAAA 2032
Db 57 AATTAAGAAAAA 37

RESULT 13
; Sequence 5796, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5796
; LENGTH: 260
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 59-LIB34-047-Q1-E1-G12
US-09-983-965-5796
```

Query Match 1.0%; Score 21; DB 10; Length 260;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATAGAGAAAAA 2031
Db 207 AATAGAGAAAAA 227

RESULT 14

US-09-960-352-4352/c
Sequence 4352, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathalagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 4352
LENGTH: 280
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 19-LIB3058-003-Q1-K1-E3
US-09-960-352-4352

Query Match 1.0%; Score 21; DB 10; Length 280;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAGAAAAA 2032
Db 49 AATAGAGAAAAA 29

RESULT 15

US-09-954-531-265/c
Sequence 265, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
FILE REFERENCE: 689290-77
CURRENT APPLICATION NUMBER: US/09/954,531
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/60/233,133
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,034
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,509
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US/60/234,567
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 1392
SOFTWARE: PatentIn version 3.0
SEQ ID NO 265
LENGTH: 411
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-531-265

Query Match 1.0%; Score 21; DB 9; Length 411;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAGAAAAA 2032
Db 44 AATAGAGAAAAA 24

RESULT 16

US-09-880-107-1438/c
Sequence 1438, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1438
LENGTH: 411
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA598749
US-09-880-107-1438

Query Match 1.0%; Score 21; DB 10; Length 411;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAGAGAAAAA 2032
Db 44 AATAGAGAAAAA 24

RESULT 17

US-09-770-445-859/c
Sequence 859, Application US/09770445
Patent No. US2002023281A1
GENERAL INFORMATION:
APPLICANT: Gerlach, Jorn
APPLICANT: An, Yong-Olang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kricker, Maya
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2023US (PARA-012PRV)
CURRENT APPLICATION NUMBER: US/09/770,445
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/178,472
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999

Query Match 1.0%; Score 21; DB 9; Length 411;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 859
; LENGTH: 785
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-859

Query Match
Best Local Similarity 100.0%; Score 21; DB 10; Length 785;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAA 2032
DB 31 AATAAGAAAAA 11

RESULT 18
US-09-770-445-263
; Sequence 263, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jörn
; APPLICANT: An, Jong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 263
; LENGTH: 986
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(986)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-445-263

Query Match
Best Local Similarity 100.0%; Score 21; DB 10; Length 986;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAA 2032
DB 158 AATAAGAAAAA 178

RESULT 19
US-09-925-301-297
; Sequence 297, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 297
; LENGTH: 998
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-297

Query Match
Best Local Similarity 100.0%; Score 21; DB 10; Length 998;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2012 AATAAGAAAAA 2032
DB 944 AATAAGAAAAA 964

RESULT 20
US-10-016-157A-89
; Sequence 89, Application US/10016157A
; Publication No. US20020192220A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yonggang
; APPLICANT: Recipon, Heve
; APPLICANT: Ghosh, Malayika
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and
; FILE REFERENCE: DEX-0253
; CURRENT APPLICATION NUMBER: US/10/016,157A
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: US 60/244,717
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 250
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 89
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-157A-89

Query Match
Best Local Similarity 100.0%; Score 21; DB 9; Length 1263;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2011 AATAAGAAAAA 2031
DB 1243 AATAAGAAAAA 1263

RESULT 21
US-09-998-598-352
; Sequence 352, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.361
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606

```

```
SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 352
; LENGTH: 1340
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-352
```

```
Query Match
Best Local Similarity 100.0%; Score 21; DB 10; Length 1340;
Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2012 AATAAGAAAAA 2032
Db 1316 AATAAGAAAAA 1336
```

```
RESULT 22
US-09-919-770-1
; Sequence 1, Application US/09919770
; Patent No. US20020048577A1
; GENERAL INFORMATION:
; APPLICANT: Bornstein, Paul
; APPLICANT: Kyriakides, Themis
; APPLICANT: Ratner, Buddy
; APPLICANT: Giachelli, Cecilia
; APPLICANT: Martinson, Laura
; APPLICANT: Scatena, Maria
; TITLE OF INVENTION: Methods and Devices to Modulate the Wound Response
; FILE REFERENCE: USFW17618
; CURRENT APPLICATION NUMBER: US/09/919,770
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,071
; PRIOR FILING DATE: 2000-08-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (102)..(1001)
US-09-919-770-1
```

```
Query Match
Best Local Similarity 100.0%; Score 21; DB 10; Length 1469;
Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2012 AATAAGAAAAA 2032
Db 1435 AATAAGAAAAA 1455
```

```
RESULT 23
US-09-925-301-259
; Sequence 259, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 259
; LENGTH: 1792
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (487)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1306)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-259
```

```
Query Match
Best Local Similarity 100.0%; Score 21; DB 10; Length 1792;
Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2012 AATAAGAAAAA 2032
Db 1768 AATAAGAAAAA 1768
```

```
RESULT 24
US-09-938-842A-4722
; Sequence 4722, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krebs, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300.3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4722
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4722
```

```
Query Match
Best Local Similarity 100.0%; Score 21; DB 9; Length 2000;
Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2012 AATAAGAAAAA 2032
Db 1735 AATAAGAAAAA 1755
```

```
RESULT 25
US-09-925-301-572
; Sequence 572, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
```

SEQ ID NO 572
LENGTH: 2013
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-301-572

Query Match
Best Local Similarity 100.0%; Score 21; DB 10; Length 2013;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2012 ATAGAGAAAAA 2032
Db 1993 ATAGAGAAAAA 2013

RESULT 26
US-09-983-965-3358/c
; Sequence 3358, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 3358
; LENGTH: 101
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 33-LIB3058-031-Q1-K1-A2
US-09-983-965-3358

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 101;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2013 ATAGAGAAAAA 2032
Db 37 ATAGAGAAAAA 18

RESULT 27
US-09-867-701-9309
; Sequence 9309, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jones, Robert
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9309
; LENGTH: 161
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9309

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 161;

Best Local Similarity 100.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2013 ATAGAGAAAAA 2032
Db 111 ATAGAGAAAAA 130

RESULT 28
US-09-983-965-2866/c
; Sequence 2866, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2866
; LENGTH: 176
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 17-LIB3058-004-Q1-K1-E1
US-09-983-965-2866

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 176;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2013 ATAGAGAAAAA 2032
Db 44 ATAGAGAAAAA 25

RESULT 29
US-09-960-352-8131/c
; Sequence 8131, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; PRIOR FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8131
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 35-LIB3057-010-Q1-K1-A12
US-09-960-352-8131

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 326;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2013 ATAGAGAAAAA 2032
Db 42 ATAGAGAAAAA 23


```
RESULT 30
US-09-960-352-5062/C
; Sequence 5062, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: MUSCLE ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5062
; LENGTH: 394
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 22-LIB3058-008-Q1-K1-F5
US-09-960-352-5062

Query Match          1.0%; Score 20; DB 10; Length 394;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
Db 44 ATAGAGAAAAA 25

RESULT 31
US-09-960-352-11099/C
; Sequence 11099, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11099
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-D8
US-09-960-352-11099

Query Match          1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 GATGCCATCTTGCTCAT 248
Db 240 GATGCCATCTTGCTCAT 221

RESULT 32
US-09-960-352-7945/C
; Sequence 7945, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
```

```
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7945
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 34-LIB3058-053-Q1-K1-A6
US-09-960-352-7945

Query Match          1.0%; Score 20; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
Db 41 ATAGAGAAAAA 22

RESULT 33
US-09-919-580-294/C
; Sequence 294, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 294
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 168..178, 195, 288, 291, 305..401
; OTHER INFORMATION: n = A,T,C or G
US-09-919-580-294

Query Match          1.0%; Score 20; DB 10; Length 410;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
Db 28 ATAGAGAAAAA 9

RESULT 34
US-09-960-352-9747/C
; Sequence 9747, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 9747
; LENGTH: 430
```

TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 42-LIB3057-018-Q1-K1-C6
US-09-960-352-9747

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 430;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 400 ATAGAGAAAAA 381

RESULT 35
US-09-834-975-616
Sequence 616, Application US/09834975
Patent No. US20020110815A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Brown, Jeffrey
APPLICANT: Bolt, Andrew
APPLICANT: Van Hufel, Christophe
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
FILE REFERENCE: MRI-016B
CURRENT APPLICATION NUMBER: US/09/834,975
CURRENT FILING DATE: 2001-04-13
PRIOR APPLICATION NUMBER: 60/197,538
PRIOR FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 1046
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 616
LENGTH: 474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)..(474)
OTHER INFORMATION: n = A,T,C or G
US-09-834-975-616

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 474;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 214 ATAGAGAAAAA 233

RESULT 36
US-09-938-842A-3890/C
Sequence 3890, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kieps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 3890
LENGTH: 477
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-3890

Query Match
Best Local Similarity 100.0%; Score 20; DB 9; Length 477;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 85 ATAGAGAAAAA 66

RESULT 37
US-09-764-846-16
Sequence 16, Application US/09764846
Patent No. US20020102638A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT212
CURRENT APPLICATION NUMBER: US/09/764,846
CURRENT FILING DATE: 2001-01-17
PRIOR APPLICATION data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 348
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 16
LENGTH: 498
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-846-16

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 498;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 460 ATAGAGAAAAA 479

RESULT 38
US-09-917-800A-306/C
Sequence 306, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendrick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
TITLE OF INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5038-US
CURRENT APPLICATION NUMBER: US/09/917,800A
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884

```

; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 306
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1,AA924767
US-09-917-800A-306
```

```

Query Match          1.0%; Score 20; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2013 ATAGAGAAAAA 2032
Db 22 ATAGAGAAAAA 3
```

```

RESULT 39
US-09-919-580-462/C
; Sequence 462, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121,552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 462
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 38, 169, 229, 278, 289, 292, 312, 329, 340, 368, 374, 414,
; LOCATION: 513, 537
; OTHER INFORMATION: n = A,T,C or G
US-09-919-580-462
```

```

Query Match          1.0%; Score 20; DB 10; Length 554;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2013 ATAGAGAAAAA 2032
Db 29 ATAGAGAAAAA 10
```

```

RESULT 40
US-09-867-550-1485/C
; Sequence 1485, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehrtan, Fuad
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60//208,427
```

```

; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1485
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-550-1485
```

```

Query Match          1.0%; Score 20; DB 10; Length 615;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2013 ATAGAGAAAAA 2032
Db 22 ATAGAGAAAAA 3
```

```

RESULT 41
US-09-925-300-548
; Sequence 548, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 548
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (61)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-548
```

```

Query Match          1.0%; Score 20; DB 10; Length 630;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2013 ATAGAGAAAAA 2032
Db 577 ATAGAGAAAAA 596
```

```

RESULT 42
US-09-925-302-240
; Sequence 240, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1998-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 240
; LENGTH: 1117
; TYPE: DNA
```

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1113)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-240

Query Match 1.0%; Score 20; DB 10; Length 1117;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 1088 ATAGAGAAAAA 1107

RESULT 43
US-09-820-893-14
Sequence 14, Application US/09820893
Patent No. US20020076705A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 31 Human Secreted Proteins
FILE REFERENCE: P2033P1
CURRENT APPLICATION NUMBER: US/09/820,893
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/531,119
PRIOR FILING DATE: 2000-03-20
PRIOR APPLICATION NUMBER: 60/102,895
PRIOR FILING DATE: 1998-10-02
NUMBER OF SEQ ID NOS: 140
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 1389
TYPE: DNA
ORGANISM: Homo sapiens
US-09-820-893-14

Query Match 1.0%; Score 20; DB 10; Length 1389;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 1347 ATAGAGAAAAA 1366

RESULT 44
US-09-938-842A-4048
Sequence 4048, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SAME, AND METHODS OF USE
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 4048
LENGTH: 1965
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-4048

Query Match 1.0%; Score 20; DB 9; Length 1965;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 1538 ATAGAGAAAAA 1557

RESULT 45
US-10-041-007-27
Sequence 27, Application US/10041007
Patent No. US20020164736A1
GENERAL INFORMATION:
APPLICANT: Matsuda, Seichi P.T.
TITLE OF INVENTION: Ginkgo Biloba Levopimaradiene Synthase
FILE REFERENCE: P02081US1
CURRENT APPLICATION NUMBER: US/10/041,007
CURRENT FILING DATE: 2002-01-07
PRIOR APPLICATION NUMBER: US 60/259,881
PRIOR FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.1
SEQ ID NO 27
LENGTH: 2429
TYPE: DNA
ORGANISM: Abies grandis
US-10-041-007-27

Query Match 1.0%; Score 20; DB 9; Length 2429;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032
DB 2407 ATAGAGAAAAA 2426

RESULT 46
US-09-822-830A-422
Sequence 422, Application US/09822830A
Patent No. US20020142952A1
GENERAL INFORMATION:
APPLICANT: Genetics Institute, Inc.
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakari
APPLICANT: Graham, James R.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6402
CURRENT APPLICATION NUMBER: US/09/822,830A
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,604
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 631
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 422
LENGTH: 2441
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-830A-422

Query Match 1.0%; Score 20; DB 10; Length 2441;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2013 ATAGAGAAAAA 2032

Db 1104 ATAGAAAAAAAAAAAA 1123

```

RESULT 47
US-09-529-063-78
; Sequence 78: Application US/09529063
; Patent No. US20020102542A1
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAIRICHI
; APPLICANT: SHIBAYAMA, SHIRO
; APPLICANT: TADA, HIDEAKI
; TITLE OF INVENTION: POLYPEPTIDE, CDNA ENCODING THE POLYPEPTIDE, AND USE OF
; TITLE OF INVENTION: THE BOTH
; FILE REFERENCE: Q58769
; CURRENT APPLICATION NUMBER: US/09/529,063
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: PCT/JP98/04514
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: JP 9-274674
; PRIOR FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 3492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (6)..(53)
; NAME/KEY: mat_peptide
; LOCATION: (54)..(2117)
; NAME/KEY: CDS
; LOCATION: (6)..(2117)
US-09-529-063-78
    
```

```

Query Match 1.0%; Score 20; DB 10; Length 3492;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2013 ATAGAAAAAAAAAAAA 2032
Db 3470 ATAGAAAAAAAAAAAA 3489
    
```

Search completed: January 10, 2003, 21:43:58
 Job time : 331 secs

This page blank (optional)

QY 121 SSIFQWENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPFEVYKACRSYSHVYLKEVR 180
 CC and/or a beta1,3gnt antisenase nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LSST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents human LSST-2.

QY 121 SSIFQWENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPFEVYKACRSYSHVYLKEVR 180
 Db 121 SSIFQWENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPFEVYKACRSYSHVYLKEVR 180
 QY 181 FENLQSLVYLLKDPNLNLIHVLVDRPRAVFRSRETKGDLIDSRIYMGHEQKLKED 240
 CC 181 FENLQSLVYLLKDPNLNLIHVLVDRPRAVFRSRETKGDLIDSRIYMGHEQKLKED 240
 Db 181 FENLQSLVYLLKDPNLNLIHVLVDRPRAVFRSRETKGDLIDSRIYMGHEQKLKED 240
 QY 241 QPYVWQVICSQDELYKTIOSLPKALQERYLLVREDLARAFAVQTSRMTEFVGLFELP 300
 CC 241 QPYVWQVICSQDELYKTIOSLPKALQERYLLVREDLARAFAVQTSRMTEFVGLFELP 300
 Db 241 QPYVWQVICSQDELYKTIOSLPKALQERYLLVREDLARAFAVQTSRMTEFVGLFELP 300
 QY 301 HQTWVHNTRFGKMGDAFHTNARDALNVSOAWRWSLPEYKVSRLQACGDAMNLLGYR 360
 CC 301 HQTWVHNTRFGKMGDAFHTNARDALNVSOAWRWSLPEYKVSRLQACGDAMNLLGYR 360
 Db 301 HQTWVHNTRFGKMGDAFHTNARDALNVSOAWRWSLPEYKVSRLQACGDAMNLLGYR 360
 QY 361 HVRSQEQGNLLDLSTWVPEQIH 386
 CC 361 HVRSQEQGNLLDLSTWVPEQIH 386
 Db 361 HVRSQEQGNLLDLSTWVPEQIH 386

RESULT 2
 ID AA011274 standard; Protein: 380 AA.
 XX AA011274:
 AC 12-MAR-2002 (first entry)
 DT 12-MAR-2002 (first entry)
 XX 12-MAR-2002 (first entry)
 DE Human L-selectin sulfotransferase-2 (LSST-2) protein.
 XX Human L-selectin sulfotransferase-2 (LSST-2) protein.
 KW Human; beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuk;
 KW antinflammatory; antiproliferative; antidiabetic; dermatological;
 KW antileukemic.
 XX Homo sapiens.
 OS Homo sapiens.
 PN WO200185177-A1.
 PD 15-NOV-2001.
 XX 15-NOV-2001.
 PF 10-MAY-2001; 2001WO-US15452.
 XX 10-MAY-2001; 2001WO-US15452.
 PR 11-MAY-2000; 2000US-0569320.
 XX 11-MAY-2000; 2000US-0569320.
 PA (BURN-) BURNHAM INST.
 XX (BURN-) BURNHAM INST.
 PI Fukuda M, Yeh J, Hiraoka N;
 XX Fukuda M, Yeh J, Hiraoka N;
 DR WPI; 2002-075226/10.
 DR N-PSDB; AAS16947.
 XX WPI; 2002-075226/10.
 PT New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 PT intestinal glucan 6-sulfotransferase
 PS Claim 21; Fig 4; 98pp; English.
 XX Claim 21; Fig 4; 98pp; English.
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active
 CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
 CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by

CC administering antibody material that specifically binds beta1,3gnt,
 CC and/or a beta1,3gnt antisenase nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LSST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents human LSST-2.

QY 7 MKLLFLVSQMAIILAFPHMYSHNISLSKKAQPERHNYLVLSWRSQSSFEVQLFGQHP 66
 CC 7 MKLLFLVSQMAIILAFPHMYSHNISLSKKAQPERHNYLVLSWRSQSSFEVQLFGQHP 66
 Db 1 MKLLFLVSQMAIILAFPHMYSHNISLSKKAQPERHNYLVLSWRSQSSFEVQLFGQHP 60
 QY 67 DVFYLMPEPAWHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEBQPRQSSLFGM 126
 CC 67 DVFYLMPEPAWHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEBQPRQSSLFGM 126
 Db 61 DVFYLMPEPAWHVMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEBQPRQSSLFGM 120
 QY 127 ENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPFEVYKACRSYSHVYLKEVRFNLOS 186
 CC 127 ENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPFEVYKACRSYSHVYLKEVRFNLOS 186
 Db 121 ENSRALCSAPACDIIPQDEILIPRAHCRLLCSQOPFEVYKACRSYSHVYLKEVRFNLOS 180
 QY 187 LYPILKDPNLNLIHVLVDRPRAVFRSRETKGDLIDSRIYMGHEQKLKEDQPYVW 246
 CC 187 LYPILKDPNLNLIHVLVDRPRAVFRSRETKGDLIDSRIYMGHEQKLKEDQPYVW 246
 Db 181 LYPILKDPNLNLIHVLVDRPRAVFRSRETKGDLIDSRIYMGHEQKLKEDQPYVW 240
 QY 247 QVICSQDELYKTIOSLPKALQERYLLVREDLARAFAVQTSRMTEFVGLFELPFIQTWV 306
 CC 247 QVICSQDELYKTIOSLPKALQERYLLVREDLARAFAVQTSRMTEFVGLFELPFIQTWV 306
 Db 241 QVICSQDELYKTIOSLPKALQERYLLVREDLARAFAVQTSRMTEFVGLFELPFIQTWV 300
 QY 307 HNITRGKMGDAFHTNARDALNVSOAWRWSLPEYKVSRLQACGDAMNLLGYRHRSQ 366
 CC 307 HNITRGKMGDAFHTNARDALNVSOAWRWSLPEYKVSRLQACGDAMNLLGYRHRSQ 366
 Db 301 HNITRGKMGDAFHTNARDALNVSOAWRWSLPEYKVSRLQACGDAMNLLGYRHRSQ 360
 QY 367 EQRNLLDLSTWVPEQIH 386
 CC 367 EQRNLLDLSTWVPEQIH 386
 Db 361 EQRNLLDLSTWVPEQIH 380

RESULT 3
 ID AAM93309 standard; Protein: 386 AA.
 XX AAM93309:
 AC AAM93309:
 DT 06-NOV-2001 (first entry)
 DT 06-NOV-2001 (first entry)
 XX 06-NOV-2001 (first entry)
 DE Human polypeptide, SFO ID NO: 2817.
 XX Human polypeptide, SFO ID NO: 2817.
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 KW Homo sapiens.
 XX Homo sapiens.
 PN EPI130094-A2.
 PD 05-SEP-2001.
 XX 05-SEP-2001.
 PF 07-JUL-2000; 2000EP-0114089.
 XX 07-JUL-2000; 2000EP-0114089.
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX 02-MAY-2000; 2000JP-0183765.
 PA (HELI-) HELIX RES INST.
 XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 DR WPI: 2001-524255/58.
 DR N-PSDB; AAK94229.

XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -

XX Claim 8: SEQ ID NO 2817; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 386 AA:

Query Match 73.8%; Score 285; DB 22; Length 386;
 Best Local Similarity 99.7%; Pred No. 5.3e-277;
 Matches 385; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSQMAILALFFHMYSHNISLSKRAQPERHMYLVLSMRSGSEFVGQ 60
 DB 1 MLPPKMKLLFLVSQMAILALFFHMYSHNISLSKRAQPERHMYLVLSMRSGSEFVGQ 60
 QY 61 LFGQHDVFIYIMPPAHVWMTFKOSTAMMLHMAVRDLIAVFLCDMSVDAYMEPPRRQ 120
 DB 61 LFGQHDVFIYIMPPAHVWMTFKOSTAMMLHMAVRDLIAVFLCDMSVDAYMEPPRRQ 120
 QY 121 SSLFQWENSRLALCSAPACDIIPQDEIIPRAHGRLLCSQQPFVEVEKACSYSHVYLKEVR 180
 DB 121 SSLFQWENSRLALCSAPACDIIPQDEIIPRAHGRLLCSQQPFVEVEKACSYSHVYLKEVR 180
 QY 181 FFMLOSLYPLKPKPSLNHIVHVRPRAVFRSRETKGDLMTDSITVNGOHOKKKED 240
 DB 181 FFMLOSLYPLKPKPSLNHIVHVRPRAVFRSRETKGDLMTDSITVNGOHOKKKED 240
 QY 241 QPYVWQVICOQSOLEIYKTIQSLPKALOERYLLVREYEDLARAVAOISMYEFGLEFLP 300
 DB 241 QPYVWQVICOQSOLEIYKTIQSLPKALOERYLLVREYEDLARAVAOISMYEFGLEFLP 300
 QY 301 HLOTWVHNITRGKGDHAFHTNARDALNVSOAMRSLPEYKVSRIQKACGDMNLGYR 360
 DB 301 HLOTWVHNITRGKGDHAFHTNARDALNVSOAMRSLPEYKVSRIQKACGDMNLGYR 360
 QY 361 HVRSEQORRLILDLSTWTVPEQIH 386
 DB 361 HVRSEQORRLILDLSTWTVPEQIH 386

RESULT 4

AA79219
 ID AAY79219 standard; Protein; 386 AA.

XX AAY79219;

DT 19-JUN-2000 (first entry)

XX Human transferase TRNSFS-11.

XX Transferrase; TRNSFS-11; human; antitumor; cell proliferation;
 KM gastrointestinal disorder; developmental disorder;
 KM genetic disorder; neurological disorder; reproductive disorder;
 KM smooth muscle disorder; immunological disorder; inflammation;

KM diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.
 XX Homo sapiens.

| Key | Location/Qualifiers |
|-------------------|--------------------------------------|
| Modified-site 121 | /note= "potential O-phosphorylation" |
| Modified-site 107 | /note= "potential O-phosphorylation" |
| Modified-site 217 | /note= "potential O-phosphorylation" |
| Modified-site 252 | /note= "potential O-phosphorylation" |
| Modified-site 364 | /note= "potential O-phosphorylation" |
| Modified-site 380 | /note= "potential O-phosphorylation" |
| Modified-site 35 | /note= "potential O-phosphorylation" |
| Modified-site 50 | /note= "potential O-phosphorylation" |
| Modified-site 81 | /note= "potential O-phosphorylation" |
| Modified-site 287 | /note= "potential O-phosphorylation" |
| Modified-site 243 | /note= "potential O-phosphorylation" |
| Modified-site 30 | /note= "potential O-phosphorylation" |
| Modified-site 308 | /note= "potential N-glycosylation" |
| Modified-site 329 | /note= "potential N-glycosylation" |
| Domain 7..23 | /note= "potential N-glycosylation" |
| Domain | /note= "transmembrane domain" |

WO200014251-A2.
 16-MAR-2000.
 09-SEP-1999; 99WO-US20989.
 10-SEP-1998; 98US-0150657.
 04-NOV-1998; 98US-0186779.
 11-MAY-1999; 99US-0133642.

(INCYTE) INCYTE PHARM INC.

Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
 Hillman JL, Azimzai Y;

WPI: 2000-256996/22.
 DR N-PSDB; AA294211.

Human transferase proteins useful for preventing, diagnosing and
 PT treating cancers and developmental, gastrointestinal, genetic,
 PT immunological, neurological, reproductive and smooth muscle disorders -

Claim 1; Page 90-91; 113pp; English.

XX The present sequence is that of human transferase TRNSFS-11, 1 of
 CC 15 claimed novel human transferase proteins of the invention (see
 CC AAY79209-23). The sequence was deduced from a cDNA clone (see
 CC AA294211). Isolated from a gallbladder library. It shows homology to
 CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is
 CC expressed in dermatologic and gastrointestinal tissues,
 CC especially those associated with inflammation and cell
 CC proliferation. The new human transferases and polynucleotides can
 CC be used in the diagnosis, prevention and treatment of cancer,
 CC developmental disorders, gastrointestinal disorders, genetic
 CC disorders, immunological disorders, neurological disorders,
 CC reproductive disorders, and smooth muscle disorders. The
 CC polypeptides can also be used to raise antibodies, and to screen

CC for agonists and antagonists of transferase activity.
 XX Sequence 386 AA:
 SQ Query Match 57.5%; Score 222; DB 21; Length 386;
 Best Local Similarity 100.0%; Pred. No. 7.7e-214;
 Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 EKACRSYSHVVLKEVRFENLQSLYPLKDPSTNLHIVHLVPRPRAVFRSRETKGDLMD 224
 |||
 DB 165 EKACRSYSHVVLKEVRFENLQSLYPLKDPSTNLHIVHLVPRPRAVFRSRETKGDLMD 224
 |||

QY 225 SRVMOQHOKLKEKEDQPYVMOVICOQSOLEYTKTIQSLPKALQERYLLVREDIARAPV 284
 |||
 DB 225 SRVMOQHOKLKEKEDQPYVMOVICOQSOLEYTKTIQSLPKALQERYLLVREDIARAPV 284
 |||

QY 285 AQSRYMEYFVGLFELPLDTWVHNITRGKMGDHAFTNARALNVSOAMRSLPYEKVS 344
 |||
 DB 285 AQSRYMEYFVGLFELPLDTWVHNITRGKMGDHAFTNARALNVSOAMRSLPYEKVS 344
 |||

QY 345 RLOKACGDAMNLGVRHVSSEQDRNLIDLDTSTVPEQIH 386
 |||
 DB 345 RLOKACGDAMNLGVRHVSSEQDRNLIDLDTSTVPEQIH 386
 |||

RESULT 5
 AAB41947
 ID AAB41947 standard; Protein; 418 AA.
 XX
 AC AAB41947;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.
 XX
 KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vulnerability; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antineuritic; antihypertensive;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;
 KW thrombosis; contraceptive.
 KW
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US086621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 XX
 PR 02-APR-1999; 99US-0127636.
 XX
 PR 05-APR-1999; 99US-0127728.
 XX
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CUBA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 XX
 DR N-PSDB: AAC76156.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 2599-2600; 5507pp; English.
 PS
 CC AACT74446 to AACT77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnerability;
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 CC anticonvulsant; osteopathic; antidiabetic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antiviral; antineuritic; antihypertensive;
 CC antianaemic; antineuritic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 SQ Sequence 418 AA;
 XX

Query Match 26.7%; Score 103; DB 21; Length 418;
 Best Local Similarity 100.0%; Pred. No. 1.7e-94;
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 VFDAVMEPPRROSSLFOWENSRALCSAPACDIIPQDEIIPRAHCRILCSQGFVEVERA 167
 |||
 DB 134 VFDAVMEPPRROSSLFOWENSRALCSAPACDIIPQDEIIPRAHCRILCSQGFVEVERA 193
 |||

QY 168 CRYSYSHVVLKEVRFENLQSLYPLKDPSTNLHIVHLVPRPRAV 210
 |||
 DB 194 CRYSYSHVVLKEVRFENLQSLYPLKDPSTNLHIVHLVPRPRAV 236
 |||

RESULT 6
 ABB81560
 ID ABB81560 standard; Protein; 171 AA.
 XX
 AC ABB81560;
 XX
 DT 05-SEP-2002 (first entry)
 XX
 DE Human high endothelial cell GLCNAC6ST portion SEQ ID NO:8.
 XX
 KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GLCNAC6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 KW
 XX
 OS Homo sapiens.
 XX
 PN US2002061562-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 09-AUG-2001; 2001US-0927602.
 XX
 PR 11-AUG-2000; 2000US-325773P.
 XX
 PA (FUKU/) FUKUDA M N.
 XX
 PA (AKAM/) AKAMA T O.
 XX
 PI Fukuda MN, Akama TO;
 XX
 DR WPI: 2002-507643/54.
 XX
 PT New nucleic acid encoding corneal
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,

PT monitoring and diagnosis of macular corneal dystrophy -
XX
XX Example 2; Fig 3c; 69pp; English.
PS
XX The present invention describes human corneal
CC N-acetylglicosamine-6-sulfotransferase (I), which is able to catalyze
CC sulfation of keratan sulfate (KS). Also described is a method for
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
CC and has ophthalmological activity. (I) can be used to treat or prevent
CC macular corneal dystrophy types I or II. (I) makes possible treatment
CC of MCD without requiring keratinoplasty or keratectomy. The present
CC sequence represents a portion of the human high endothelial cell
CC N-acetylglicosamine-6-sulfotransferase (GlcNAc6ST), which is given in
CC the exemplification of the present invention.
XX
SQ Sequence 171 AA;
Query Match 15.5%; Score 60; DB 23; Length 171;
Best Local Similarity 100.0%; Pred. No. 9.5e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 166 KACRSYSHVVLKEVRFNNQSLYPLKDPSLNLIHVLDPRVFRSRERTKGDLIMIDS 225
DB 53 KACRSYSHVVLKEVRFNNQSLYPLKDPSLNLIHVLDPRVFRSRERTKGDLIMIDS 112
RESULT 7
AA39919
ID AA39919 standard; Protein: 388 AA.
XX
AC AA39919;
XX
DT 08-DEC-1999 (first entry)
XX
DE Mouse glycosyl sulfotransferase-3 protein sequence.
XX
XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KM selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ.
XX
XX Mus sp.
OS
XX
PN WO949018-A1.
XX
PD 30-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
XX
XX (REGC) UNIV CALIFORNIA.
PA (SYNT) SYMTEX USA INC.
XX
PI Bistup A, Rosen SD, Tangemann K, Hemmerich S;
XX
XX WPI; 1999-580442/49.
DR N-PSDB; AA220793.
XX
PT Human and murine glycosyl sulfotransferase 3 and related
PT polynucleotides -
XX
PS Claim 2; Fig 4; 59pp; English.
XX
XX This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.

XX
SQ Sequence 388 AA;
Query Match 10.1%; Score 39; DB 20; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 44 HVLVLSMRSGSSFFVGOJFGQHPDYFLMEPPAHYWMTF 82
DB 43 HVLVLSMRSGSSFFVGOJFGQHPDYFLMEPPAHYWMTF 81
RESULT 8
ABB81558
ID ABB81558 standard; Protein: 169 AA.
XX
AC ABB81558;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human corneal N-acetylglicosamine-6-sulfotransferase portion SEQ.6.
XX
XX Human; N-acetylglicosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KM corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
XX Homo sapiens.
OS
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PF 09-AUG-2001; 2001US-0927602.
XX
PR 11-AUG-2000; 2000US-325773P.
XX
XX (FUKU/) FUKUDA M N.
PA (AKAW/) AKAMA T O.
XX
PI Fukuda MN, Akama TO;
XX
DR WPI; 2002-507643/54.
XX
PT New nucleic acid encoding corneal
PT N-acetylglicosamine-6-sulfotransferase, useful for treatment,
PT monitoring and diagnosis of macular corneal dystrophy -
XX
XX Example 2; Fig 3c; 69pp; English.
PS
XX
XX The present invention describes human corneal
CC N-acetylglicosamine-6-sulfotransferase (II), which is able to catalyze
CC sulfation of keratan sulfate (KS). Also described is a method for
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC and detecting susceptibility to MCD. (II) is located to chromosome 16q22,
CC and has ophthalmological activity. (II) can be used to treat or prevent
CC macular corneal dystrophy types I or II. (II) makes possible treatment
CC of MCD without requiring keratinoplasty or keratectomy. The present
CC sequence represents a portion of the human corneal N-acetylglicosamine-
CC 6-sulfotransferase (GlcNAc6ST), which is given in the exemplification
CC of the present invention.
XX
SQ Sequence 169 AA;
Query Match 4.9%; Score 19; DB 23; Length 169;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 44 HVLVLSMRSGSSFFVGOJFGQHPDYFLMEPPAHYWMTF 62
DB 10 HVLVLSMRSGSSFFVGOJFGQHPDYFLMEPPAHYWMTF 28
RESULT 9

ABB81559
ID ABB81559 standard; Protein: 169 AA.
XX
AC ABB81559;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human intestinal N-acetylglucosamine-6-sulfotransferase portion SEQ:7.
XX
KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GICNA6ST;
KW Corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW Ophthalmological.
XX
OS Homo sapiens.
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PE 09-AUG-2001; 2001US-0927602.
XX
PR 11-AUG-2000; 2000US-325773P.
XX
PA (EUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
XX
PI Fukuda MN, Akama TO;
XX
DR WPI; 2002-507643/54.
XX
PT New nucleic acid encoding corneal
PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
PT monitoring and diagnosis of macular corneal dystrophy
XX
PS Example 2; Fig 3C; 69pp; English.
XX
CC The present invention describes human corneal
CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze
CC sulfation of keratan sulfate (KS). Also described is a method for
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
CC and has ophthalmological activity. (I) can be used to treat or prevent
CC macular corneal dystrophy types I or II. (I) makes possible treatment
CC of MCD without requiring keratoplasty or keratectomy. The present
CC sequence represents a portion of the human intestinal
CC N-acetylglucosamine-6-sulfotransferase (GICNA6ST), which is given in
CC the exemplification of the present invention.
XX
SQ Sequence 169 AA;
XX

Query Match 4.9%; Score 19; DB 23; Length 169;
Best Local Similarity 100.0%; Pred. No. 1,2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 167 ACRSYSHVLEKVEFFNLQ 185
DB 51 ACRSYSHVLEKVEFFNLQ 69

RESULT 10
AAV72639
ID AAV72639 standard; Protein: 390 AA.
XX
AC AAV72639;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-4alpha (GST-4alpha).
XX
KW Human: glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW

glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1.
XX
OS Homo sapiens.
XX
PN WO200106015-A1.
XX
PD 25-JAN-2001.
XX
PE 19-JUL-2000; 2000WO-US19741.
XX
PR 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Rosen SD, Lee JK, Hemmerich S;
XX
DR WPI; 2001-138471/14.
DR N-PSDB; AAD02697, AAD02698, AAD02699.
XX
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications
XX
PS Claim 3; Fig 1; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
CC alpha). GST-4 gene is found on chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.
XX
SQ Sequence 390 AA;
XX

Query Match 4.9%; Score 19; DB 22; Length 390;
Best Local Similarity 100.0%; Pred. No. 2,6e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 167 ACRSYSHVLEKVEFFNLQ 185
DB 165 ACRSYSHVLEKVEFFNLQ 183

RESULT 11
ABB81556
ID ABB81556 standard; Protein: 390 AA.
XX
AC ABB81556;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
XX
KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GICNA6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW

| | |
|-----------|--|
| KW | ophthalmological. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | US2002061562-A1. |
| XX | |
| PD | 23-MAY-2002. |
| XX | |
| PE | 09-AUG-2001; 2001US-0927602. |
| XX | |
| PR | 11-AUG-2000; 2000US-325773P. |
| XX | |
| PA | (FUKU/) FUKUDA M N. |
| PA | (AKAM/) AKAMA T O. |
| XX | |
| PI | Fukuda MN, Akama TO; |
| DR | WPI; 2002-507643/54. |
| XX | |
| PT | New nucleic acid encoding corneal |
| PT | N-acetylglucosamine-6-sulfotransferase, useful for treatment, |
| PT | monitoring and diagnosis of macular corneal dystrophy |
| XX | |
| PS | Example 5; Fig 2A-B; 69pp; English. |
| XX | |
| CC | The present invention describes human corneal |
| CC | N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze |
| CC | sulfation of keratan sulfate (KS). Also described is a method for |
| CC | monitoring the effect of treatments for macular corneal dystrophy (MCD), |
| CC | and detecting susceptibility to MCD. (I) is located to chromosome 16q22, |
| CC | and has optalmalogical activity. (I) can be used to treat or prevent |
| CC | macular corneal dystrophy types I or II. (II) makes possible treatment |
| CC | of MCD without requiring keratoplasty or keratectomy. The present |
| CC | sequence represents human intestinal N-acetylglucosamine-6- |
| CC | sulfotransferase, which is given in comparison with (I) in the |
| CC | exemplification of the present invention. |
| SQ | Sequence 390 AA: |
| | |
| | Query Match 4.9%; Score 19; DB 23; Length 390; |
| | Best Local Similarity 100.0%; Pred. No. 2.6e-10; |
| | Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0 |
| QY | 167 ACRSYSHVYLKEVPFNLTQ 185 |
| DB | 165 ACRSYSHVYLKEVPFNLTQ 183 |
| RESULT 12 | |
| AAY72638 | |
| ID | AAY72638 standard; Protein; 395 AA. |
| XX | |
| AC | AAY72638; |
| XX | |
| DT | 02-MAY-2001 (first entry) |
| XX | |
| DE | Mouse glycosyl sulfotransferase-4 (GST-4). |
| XX | |
| KW | Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive; |
| KW | therapy; selectin binding inhibitor; gene therapy; inflammation; |
| KW | systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; |
| KW | polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; |
| KW | glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailtis; |
| KW | Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; |
| KW | demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; |
| KW | myocarditis; adult respiratory distress syndrome; eczema; psoriasis; |
| KW | asthma; hypersensitivity; rheumatic fever; tissue rejection; |
| KW | chromosome 8E1. |
| OS | Mus musculus. |
| XX | |
| PN | WO200106015-A1. |
| XX | |

XX 25-JAN-2001.
PD
PF 19-JUL-2000; 2000WO-US19741.
PE
PR 20-JUL-1999; 99US-0144694.
PR 13-JUL-2000; 2000US-0593828.
PX
PA (REGC) UNTV CALIFORNIA.
PY
PI Rosen SD, Lee JK, Hemmerich S;
XX WPI: 2001-138471/14.
DR N-PSDB; AAD02696.
XX
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications -
XX
XX
PS Claim 3; Fig 2; 128pp; English.
XX
XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
CC gene is found on chromosome 8B1.
CC
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
CX
SQ Sequence 395 AA:

Query Match 4.9%; Score 19; DB 22; Length 395;
Best Local Similarity 100.0%; Pred. No. 2,7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 44 HVLVLSWRSGSEFVGQLF 62
|||||
DB 42 HVLVLSWRSGSSFFVGQLF 60

RESULT 13
ID AAY72640
AAY72640
AC AAY72640;
XX
DT 02-MAY-2001 (first entry)
DE Human glycosyl sulfotransferase-4beta (GST-4beta).
XX
XX Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polycystic kidney disease; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailtis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1.
OS Homo sapiens.
XX
PN WO200106015-A1.

PD 25-JAN-2001.
 XX
 PF 19-JUL-2000; 2000WO-US19741.
 XX
 XX 20-JUL-1999; 99US-0144694.
 PR 13-JUL-2000; 2000US-0593826.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Rosen SD, Lee JK, Hemmerich S;
 XX
 DR WPI: 2001-138471/14.
 DR N-PSDB: AAD02697, AAD02700.
 XX
 PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications
 XX
 PS Claim 3: Fig 4B; 128pp; English.
 XX
 CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.
 XX
 SQ Sequence 395 AA:
 XX
 Query Match 4.9%; Score 19; DB 22; Length 395;
 Best Local Similarity 100.0%; Pred. No. 2.7e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 44 HVLVSSWRSGSSFVGOLF 62
 DB 42 HVLVSSWRSGSSFVGOLF 60
 XX
 RESULT 14
 ABB81554
 ID ABB81554 standard; Protein: 395 AA.
 XX
 AC ABB81554;
 XX
 DT 05-SEP-2002 (first entry)
 XX
 DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
 XX
 XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GLCNAC6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological; chromosome 16q22.
 XX
 OS Homo sapiens.
 XX
 PN US2002061562-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 09-AUG-2001; 2001US-0927602.
 XX
 PR 11-AUG-2000; 2000US-325773P.
 XX
 PA (FUKU/) FUKUDA M N.

PA (AKAM/) AKAMA T O.
 XX
 PI Fukuda MN, Akama TO;
 XX
 DR WPI: 2002-507643/54.
 DR N-PSDB: ABB89506.
 XX
 PT New nucleic acid encoding corneal
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
 PT monitoring and diagnosis of macular corneal dystrophy
 XX
 PS Claim 13: Fig 1A-D; 69pp; English.
 XX
 CC The present sequence represents human corneal
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze
 CC sulfation of keratan sulfate (KS). Also described is a method for
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
 CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular corneal dystrophy types I or II. (I) makes possible treatment
 CC of MCD without requiring keratoplasty or keratectomy.
 XX
 SQ Sequence 395 AA:
 XX
 Query Match 4.9%; Score 19; DB 23; Length 395;
 Best Local Similarity 100.0%; Pred. No. 2.7e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 44 HVLVSSWRSGSSFVGOLF 62
 DB 42 HVLVSSWRSGSSFVGOLF 60
 XX
 RESULT 15
 ABB81555
 ID ABB81555 standard; Protein: 395 AA.
 XX
 AC ABB81555;
 XX
 DT 05-SEP-2002 (first entry)
 XX
 DE Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
 XX
 XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GLCNAC6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 XX
 OS Homo sapiens.
 OS Mus musculus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FH Misc-difference 10 /label= Ala, Thr, Val
 FT /label= Ala, Val, Ser
 FT Misc-difference 20 /label= Phe, Cys, Gly
 FT Misc-difference 39 /label= Ala, Asp, Glu
 FT Misc-difference 96 /label= Val, Met, Ile
 FT Misc-difference 142 /label= Ala, Thr, Asn
 FT Misc-difference 147 /label= Ala, Asp, Glu
 FT Misc-difference 159 /label= Thr, Ser, Gly
 FT Misc-difference 238 /label= Gly, His, Arg
 FT Misc-difference 294 /label= Ser, Thr, Gly
 FT Misc-difference 371

FT /label= Ala, Thr, Ser
 FT Misc-difference 380 /label= Leu, Pro, Met
 FT Misc-difference 382 /label= Gly, His, Ser
 FT Misc-difference 384 /label= Thr, Ser, Lys
 FT Misc-difference 390 /label= Ala, Glu
 FT Misc-difference 391 /label= Ser, Lys
 FT Misc-difference 392 /label= His, Gln
 FT Misc-difference 394 /label= Arg, Glu
 FT Misc-difference 395 /label= Asn, Ser
 FT US2002061562-A1.
 PD 23-MAY-2002.
 PF 09-AUG-2001; 2001US-0927602.
 PR 11-AUG-2000; 2000US-325773P.
 PA (FUKU/) FUKUDA M N.
 PA (AKAMA/) AKAMA T O.
 PI Fukuda MN, Akama TO;
 PI WPI: 2002-507643/54.
 DR
 XX
 PT New nucleic acid encoding corneal
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
 PT monitoring and diagnosis of macular corneal dystrophy -
 PT
 XX
 PS Example 5; Fig 2A-B; 69pp; English.
 XX
 CC The present invention describes human corneal
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyse
 CC sulfation of keratan sulfate (KS). Also described is a method for
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
 CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular corneal dystrophy types I or II. (I) makes possible treatment
 CC of MCD without requiring keratoplasty or keratectomy. The present
 CC sequence represents a consensus N-acetylglucosamine-6-sulfotransferase
 CC which is given in the exemplification of the present invention.
 CC
 XX
 SQ Sequence 395 AA;
 Query Match 4.9%; Score 19; DB 23; Length 395;
 Best Local Similarity 100.0%; Pred. No. 2.7e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 44 HVLVLSMRSGSSFGQLF 62
 Db 42 HVLVLSMRSGSSFGQLF 60
 RESULT 16
 AAEL15438
 ID AAEL15438 standard; Protein; 395 AA.
 XX
 AC AAEL15438;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Human drug metabolizing enzyme (DME)-5.
 XX
 KW Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
 KW Inflammatory disorder; acquired immune deficiency syndrome; infection;

KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
 KW cancer; endocrine disorder; hypochalaemus disorder; pituitary disorder;
 KW gastrointestinal disorder; metabolic disorder; developmental disorder;
 KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KW DME-5.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..32
 FT /label= Signal_peptide
 FT Protein 33..395
 FT /note= "Human mature DME-5 protein"
 FT Peptide 1..35
 FT /label= Signal_peptide
 FT Protein 36..395
 FT /note= "Human mature DME-5 protein"
 FT
 FT
 PV WO200179468-A2.
 XX
 PD 25-OCT-2001.
 XX
 PD 12-APR-2001; 2001WO-0511869.
 PF
 PR 13-APR-2000; 2000US-197590P.
 PR 19-APR-2000; 2000US-198403P.
 PR 28-APR-2000; 2000US-200185P.
 PR 05-MAY-2000; 2000US-202234P.
 PR 11-MAY-2000; 2000US-203509P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Policky JL, Hatella A, Burford N, Ring HZ, Lal P, Tribouley CM;
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 PI Au-Young J;
 XX
 DR WPI: 2002-066363/09.
 DR N-PSDB: AAD24670.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anaemia, asthma,
 PT infertility -
 PT
 XX
 PS Claim 1a; Page 131-132; 143pp; English.
 XX
 CC The invention relates to human drug metabolizing enzymes referred as
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
 CC of the invention are useful for assessing toxicity of test compounds
 CC and in gene therapy. Sequences of the invention are useful in the
 CC diagnosis, prevention and treatment of autoimmune/inflammatory
 CC disorders such as acquired immune deficiency syndrome (AIDS), adult
 CC respiratory distress syndrome, allergies, anaemia, arteriosclerosis,
 CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
 CC infections; cell proliferative disorders such as actinic keratosis,
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus
 CC and pituitary resulting from lesions such as primary brain tumours,
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
 CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
 CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such

CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-5 protein.

XX Sequence 395 AA;

Query Match 4.9%; Score 19; DB 23; Length 395;

Best Local Similarity 100.0%; Pred. No. 2,7e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVSSWRSGSSFFVGOLF 62

DB 42 HVLVSSWRSGSSFFVGOLF 60

RESULT 17

AAU11275

ID AAU11275 standard; Protein; 395 AA.

XX AAU11275;

DT 12-MAR-2002 (first entry)

DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.

XX Mouse; beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; anticancer;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.

XX Mus musculus.

XX WO200185177-A1.

XX 15-NOV-2001.

XX 10-MAY-2001; 2001WO-US15452.

XX 11-MAY-2000; 2000US-0569320.

XX (BURN-) BURHAM INST.

XX Fukuda M, Yeh J, Hirooka N;

XX WPI; 2002-075226/10.

DR N-PSDB; AAS16948.

PT New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 PT intestinal GlcNAc 6-sulfotransferase

XX Claim 28; Fig 10; 98pp; English.

XX The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active
 CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
 CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3gnt,
 CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3gnt. The method is useful for treating L-selectin

CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents mouse I-GlcNAc6ST.

XX Sequence 395 AA;

Query Match 4.9%; Score 19; DB 23; Length 395;

Best Local Similarity 100.0%; Pred. No. 2,7e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVSSWRSGSSFFVGOLF 62

DB 42 HVLVSSWRSGSSFFVGOLF 60

RESULT 18

ABB81557

ID ABB81557 standard; Protein; 418 AA.

XX ABB81557;

DT 05-SEP-2002 (first entry)

DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.

XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.

XX Mus musculus.

XX US2002061562-A1.

XX 23-MAY-2002.

XX 09-AUG-2001; 2001US-0927602.

XX 11-AUG-2000; 2000US-325773P.

XX (FUKU) FUKUDA M N.

PA (AKAM/) AKAMA T O.

XX Fukuda MN, Akama TO;

XX WPI; 2002-507643/54.

PT New nucleic acid encoding corneal
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
 PT monitoring and diagnosis of macular corneal dystrophy
 XX Example 5; Page 24-25; 69pp; English.

XX The present invention describes human corneal
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze
 CC sulfation of keratan sulfate (KS). Also described is a method for
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
 CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular corneal dystrophy types I or II. (I) makes possible treatment
 CC of MCD without requiring keratoplasty or keratotomy. The present
 CC sequence represents mouse intestinal N-acetylglucosamine-6-
 CC sulfotransferase, which is given in comparison with (I) in the
 CC exemplification of the present invention.

XX Sequence 418 AA;

Query Match 4.9%; Score 19; DB 23; Length 418;

Best Local Similarity 100.0%; Pred. No. 2,8e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVSSWRSGSSFFVGOLF 62

Mon Jan 13 09:14:02 2003

us-09-816-825-2.0115.rag

Page 11

Db 66 HVLVLSWRSRSGSFVGOLF 84

Search completed: January 11, 2003, 01:29:48
Job time : 37 secs

This Page Blank (uspto)

Mon Jan 13 09:14:04 2003

us-09-816-825-2.olil5.rpr

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:29:11 ; Search time 19 seconds
(without alignments)
1953.047 Million cell updates/sec

Title: US-09-816-825-2
Perfect score: 386
Sequence: 1 MLLPKKKMLLLFLVSQMAIL.....EGRNLLDLLSTWVPEQIH 386

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Length | ID | Description |
|------------------|-------------|--------|----|-------------|
| No matches found | | | | |

Search completed: January 11, 2003, 01:31:32
Job time : 19 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:23:06 ; Search time 13 Seconds
(without alignments)
1231.528 Million cell updates/sec

Title: US-09-816-825-2
Perfect score: 386
Sequence: 1 MLIPKMKLLFLVSOMAIL.....EQRNLLDLSTWVPEQIH 386

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 segs, 41476328 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Score | Match | length | ID | Description |
|------------------|----------------|-------|--------|----|-------------|
| ----- | | | | | |
| No matches found | | | | | |

Search completed: January 11, 2003, 01:30:15
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:23:56 ; Search time 31 seconds

(without alignments)
2565.619 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 386

Sequence: 1 MLPPKMKLLFLVSQMAIL.....EQRNLLDLSTWTVPQDIH 386

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 671580 segs, 206047115 residues

Word size: 15

Total number of hits satisfying chosen parameters: 6

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database:

SPREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 386 | 100.0 | 386 | 4 Q9Y5R3 | Q9Y5R3 homo sapien |
| 2 | 386 | 10.1 | 388 | 11 Q9R1I1 | Q9R1I1 mus musculu |
| 3 | 39 | 10.1 | 388 | 11 Q9WDE5 | Q9WDE5 mus musculu |
| 4 | 19 | 4.9 | 395 | 4 Q9GZX3 | Q9GZX3 homo sapien |
| 5 | 19 | 4.9 | 395 | 11 Q9QUP4 | Q9QUP4 mus musculu |
| 6 | 19 | 4.9 | 411 | 4 Q9GZS9 | Q9GZS9 homo sapien |

ALIGNMENTS

RESULT 1
Q9Y5R3 PRELIMINARY: PRT: 386 AA.
AC Q9Y5R3:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
DE Sulfotransferase GST-3).
GN GST3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TONSIL;
RX MEDLINE=99264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TONSIL;
RX "Chromosomal Localization and Genomic Organization for the
RT Galactose/N-Acetylglucosamine/N-Acetylglucosamine 6-O-
RT Sulfotransferase Gene Family.";
RL Glycobiology 0:0-0(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21332592; PubMed=11439191;
RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Elies L.G.,
RA Rabuka D., Hindsgrau O., Marth J.D., Lowe J.B., Fakuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT core1 extension beta1,3-N-acetylglucosaminyltransferase.";
RL Cell 105:957-969(2001).
DR EMBL: AF131235; AAC33015.1; -;
DR EMBL: AF280088; AAG48245.1; -;
DR EMBL: AF149783; AAK48417.1; -;
DR InterPro: IPR001092; HLH_basic.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW Lectin; Selectin; Transferase.
SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417743A CRC64;

Query Match 100.0%; Score 386; DB 4; Length 386;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 1 | MLPPKMKLLFLVSQMAILFFHMYSHNSSLKKAQPERHAYVLSMSRSGSSFFVQ | 60 |
| DB | 1 | MLPPKMKLLFLVSQMAILALFFHMYSHNSSLKKAQPERHAYVLSMSRSGSSFFVQ | 60 |
| QY | 61 | LFQGHDPVEYLMEPAHWMTFKOSTAWMLHMAVDLIRAVFLCDMSVDFAYMEPGRQ | 120 |
| DB | 61 | LFQGHDPVEYLMEPAHWMTFKOSTAWMLHMAVDLIRAVFLCDMSVDFAYMEPGRQ | 120 |
| QY | 121 | SSLFQWENSRALCSAPACDIIPODEIIPRAHGRLLCSQPFVEVEKACSYSVVLKEYR | 180 |
| DB | 121 | SSLFQWENSRALCSAPACDIIPODEIIPRAHGRLLCSQPFVEVEKACSYSVVLKEYR | 180 |
| QY | 181 | FFNLOSLLPLKDPSSLNLHVLVDRPRAVFRSREKTKGDLMSIRIVGQHBOKLKED | 240 |
| DB | 181 | FFNLOSLLPLKDPSSLNLHVLVDRPRAVFRSREKTKGDLMSIRIVGQHBOKLKED | 240 |
| QY | 241 | QPYVAVQVTCQSOLEIYKTQSIPKALQERYLLVREDLARAVAGTSRYEFGLEFLP | 300 |
| DB | 241 | QPYVAVQVTCQSOLEIYKTQSIPKALQERYLLVREDLARAVAGTSRYEFGLEFLP | 300 |
| QY | 301 | HLQTVWNITRGKGMGDHAFHTNARDALNSQAMPWSLPEYKYSRLQKACGDAAMLGYR | 360 |
| DB | 301 | HLQTVWNITRGKGMGDHAFHTNARDALNSQAMPWSLPEYKYSRLQKACGDAAMLGYR | 360 |
| QY | 361 | HVNSEQQRNLLDLSTWTVPQDIH | 386 |

Db 361 HVHSEQFQRLNLLDLSTWTPVPEQIH 386

RESULT 2

ID 09R111 PRELIMINARY; PRT: 388 AA.

AC 09R111; 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

GN L-selectin ligand sulfotransferase.

OS Mus musculus (Mouse).

OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RA MEDLINE=99361934; PubMed=10435581;

RA Hirooka N., Petrlyak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,

12awa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;

RT A novel, high endothelial venule-specific sulfotransferase expresses

RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.;

RL Immunity 11:79-89(1999).

DR EMBL: AF109155; AAD45579.1; -

DR MGD: MGI:1349479; Chst4.

DR Lectin; Selectin; Transferase.

SC SEQUENCE 388 AA; 44636 MW; 6D5371AFB684AEE CRC64;

Query Match 10.1%; Score 39; DB 11; Length 388;

Best Local Similarity 100.0%; Pred. No. 1.4e-34;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 44 HVLVSSWRSGSFVGOLEFGQHPDVFYLMPEPAHWMTF 82

43 HVLVSSWRSGSFVGOLEFGQHPDVFYLMPEPAHWMTF 81

RESULT 3

09WTE5 PRELIMINARY; PRT: 388 AA.

AC 09WTE5; 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (Chondroitin

DE 6-Keratan) sulfotransferase 4).

GN Chst4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RA STRAIN=C57BL/6; TISSUE=TONSTL;

RA MEDLINE=99264336; PubMed=10330415;

RA Bistup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,

Huang C.-C., Kanaghi R., Rosen S.D., Hemmerich S.;

RT Sulfotransferases of two specificities function in the reconstitution

RT of high endothelial cell ligands for L-selectin.;

RL J. Cell Biol. 143:899-910(1999).

DR EMBL: AF219990; AAG26325.1; -

DR MGD: MGI:1349479; Chst4.

DR Lectin; Selectin; Transferase.

SC SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

Query Match 4.9%; Score 19; DB 4; Length 395;

Best Local Similarity 100.0%; Pred. No. 3.5e-12;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 44 HVLVSSWRSGSFVGOLEFGQHPDVFYLMPEPAHWMTF 82

42 HVLVSSWRSGSFVGOLEFGQHPDVFYLMPEPAHWMTF 60

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bona M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Holmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,

RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,

RA Hayashizaki Y.,

RT "Functional annotation of a full-length mouse cDNA collection.;"

RL Nature 409:685-690(2001).

DR EMBL: AF131236; AAD33016.1; -

DR EMBL: AK009113; BAB26078.1; -

DR MGD: MGI:1349479; Chst4.

DR Transferase.

SC SEQUENCE 388 AA; 44694 MW; D1E9D7796DE8574D CRC64;

Query Match 10.1%; Score 39; DB 11; Length 388;

Best Local Similarity 100.0%; Pred. No. 1.4e-34;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 44 HVLVSSWRSGSFVGOLEFGQHPDVFYLMPEPAHWMTF 82

43 HVLVSSWRSGSFVGOLEFGQHPDVFYLMPEPAHWMTF 81

RESULT 4

09GZX3 PRELIMINARY; PRT: 395 AA.

AC 09GZX3; 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal N-

DE acetylglucosamine-6-O-sulfotransferase).

GN GST4BETA OR ChST6.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RA Rosen S.D., Lee J.-K., Bhakta S., Bistup A., Rudde N.R.,

RT Hemmerich S.;

RT "Chromosomal Localization and Genomic Organization for the

RT Galactose/N-acetylgalactosamine/N-Acetylglucosamine 6-O-

RT Sulfotransferase Gene Family.;"

RL Glycobiology 0:0-0(2001).

RP SEQUENCE FROM N.A.

RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,

RA Nakamura T., Doi A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,

RA Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;

RT "Macular corneal dystrophy type I and type II are caused by distinct

RT mutations in a new sulfotransferase gene.;"

RL Nat. Genet. 26:237-241(2000).

DR EMBL: AF280086; AAG48244.1; -

DR EMBL: AF219990; AAG26325.1; -

DR EMBL: AF219991; AAG26327.1; -

DR Transferase.

SC SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

Query Match 4.9%; Score 19; DB 4; Length 395;

Best Local Similarity 100.0%; Pred. No. 3.5e-12;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 44 HVLVSSWRSGSFVGOLEFGQHPDVFYLMPEPAHWMTF 82

42 HVLVSSWRSGSFVGOLEFGQHPDVFYLMPEPAHWMTF 60

RESULT 5
 ID 090UP4 PRELIMINARY: PRT: 395 AA.
 AC 090UP4:
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase.
 GN CHST5 OR I-GLCNAC-6-ST.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=INTESTINE;
 RX MEDLINE=99423439; PubMed=10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
 RT Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL/ AF176841; AAD56003.1; -
 DR EMBL/ AF176840; AAD56002.1; -
 DR MGD: MGI:1931825; Chst5.
 KW Transferase.
 SQ SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64;

Query Match 4.9%; Score 19; DB 11; Length 395;
 Best Local Similarity 100.0%; Pred. No. 3.5e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 HVLVSSMRSGSFRVQGLF 62
 |||
 DB 42 HVLVSSMRSGSFRVQGLF 60

RESULT 6

09GZS9 PRELIMINARY: PRT: 411 AA.

AC 09GZS9; 09UBI3;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Intestinal GLCNAC-6-sulfotransferase (Intestinal N-
 DE acetylglucosamine-6-O-sulfotransferase) (N-acetylglucosamine
 DE 6-O-sulfotransferase).
 GN CHST5 OR I-GLCNAC-6-ST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20472330; PubMed=11017086;
 RA Akama T.O., Nishida K., Nakayama T., Watanabe H., Ozaki K.,
 RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
 RA Fujiwara T., Rhonara E.J., Shimomura Y., Kinoshita S., Tanigami A.,
 RA Fukuda M.N.;
 RT "Macular corneal dystrophy type I and type II are caused by distinct
 RT mutations in a new sulphotransferase gene.";
 RL Nat. Genet. 26:237-241(2000).
 RN [2]
 RP SEQUENCE OF 22-411 FROM N.A.
 RC TISSUE=INTESTINE;
 RX MEDLINE=99423439; PubMed=10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
 RT Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL/ AF246718; AAG28023.1; -
 DR EMBL/ AF219991; AAG26326.1; -
 DR EMBL/ AF176839; AAD56001.1; -
 DR EMBL/ AF176838; AAD56000.1; -

KW Transferase.
 SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;
 Query Match 4.9%; Score 19; DB 4; Length 411;
 Best Local Similarity 100.0%; Pred. No. 3.6e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 ACRSYSHVYLKEVRFNLTQ 185
 |||
 DB 186 ACRSYSHVYLKEVRFNLTQ 204

Search completed: January 11, 2003, 01:31:00
 Job time : 32 secs

This Page Blank (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:29:56 ; Search time 16 Seconds

(without alignments)
709.828 Million cell updates/sec

Title: US-09-816-825-2

Sequence: 1 MLLPKKKLLFLVSQMAIL.....EQNNLLDLSTWTFEQIH 386

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 15

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|------------------|
| 1 | 386 | 100.0 | 386 | 4 | US-09-045-284A-2 |
| 2 | 386 | 100.0 | 386 | 4 | US-09-190-911-1 |

ALIGNMENTS

RESULT 1
US-09-045-284A-2
Sequence 2, Application US/09045284A
Patent No. 6265192
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107051
CURRENT APPLICATION NUMBER: US/09/045, 284A
CURRENT FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 386
TYPE: PRT
ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 100.0%; Score 386; DB 4; Length 386;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | | | | | | | |
|----|-----|-------|---------------|------------|-------------|---------|---------|---------|------------------|
| QY | 1 | MLLPK | KKLLFLVSQMAIL | FFHMYSHN | SSLSKKAQPER | MHVLYLS | SMRSGSS | FFVQ | 60 |
| DB | 1 | MLLPK | KKLLFLVSQMAIL | ALFFHMYSHN | SSLSKKAQPER | MHVLYLS | SMRSGSS | FFVQ | 60 |
| QY | 61 | LFQGH | PDVFLMEPRAM | VMWTFKOST | AMLMHAYRDL | IRAVFL | CDMSV | PDAYME | PPRRQ 120 |
| DB | 61 | LFQGH | PDVFLMEPRAM | VMWTFKOST | AMLMHAYRDL | IRAVFL | CDMSV | PDAYME | PPRRQ 120 |
| QY | 121 | SSLFQ | MENSRALCSAP | ACDIIPODEI | IPRAHCR | LLCSQGF | FEVEKAC | RSYSHV | LKEVR 180 |
| DB | 121 | SSLFQ | MENSRALCSAP | ACDIIPODEI | IPRAHCR | LLCSQGF | FEVEKAC | RSYSHV | LKEVR 180 |
| QY | 181 | FFNLQ | SLYPLKDP | SLNLHIVL | VRPRAV | FRSRERT | KGDMIS | RIVMQ | HEQKLKED 240 |
| DB | 181 | FFNLQ | SLYPLKDP | SLNLHIVL | VRPRAV | FRSRERT | KGDMIS | RIVMQ | HEQKLKED 240 |
| QY | 241 | QPYVM | QVICO | SOLEIKT | QSLPKALQ | ERYLLV | REDLAR | AAVQTS | RYTEFVGLFLP 300 |
| DB | 241 | QPYVM | QVICO | SOLEIKT | QSLPKALQ | ERYLLV | REDLAR | AAVQTS | RYTEFVGLFLP 300 |
| QY | 301 | HLQTV | HNITRGK | GMGDA | FHTNARD | ALNYSQ | AMRSLPE | EKYSRLQ | AKGADAMNLGYR 360 |
| DB | 301 | HLQTV | HNITRGK | GMGDA | FHTNARD | ALNYSQ | AMRSLPE | EKYSRLQ | AKGADAMNLGYR 360 |
| QY | 361 | HVRSE | QQRNLLDL | SLTWTFEQIH | 386 | | | | |
| DB | 361 | HVRSE | QQRNLLDL | SLTWTFEQIH | 386 | | | | |

RESULT 2

US-09-190-911-1
Sequence 1, Application US/09190911
Patent No. 6365365
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Tangemann, Kirsten
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CIP
CURRENT APPLICATION NUMBER: US/09/190, 911
CURRENT FILING DATE: 1998-11-12
EARLIER APPLICATION NUMBER: 09/045, 284
EARLIER FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 386
TYPE: PRT
ORGANISM: H. sapiens
US-09-190-911-1

Query Match 100.0%; Score 386; DB 4; Length 386;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | | | | | | | |
|----|-----|-------|---------------|------------|-------------|---------|---------|--------|--------------|
| QY | 1 | MLLPK | KKLLFLVSQMAIL | ALFFHMYSHN | SSLSKKAQPER | MHVLYLS | SMRSGSS | FFVQ | 60 |
| DB | 1 | MLLPK | KKLLFLVSQMAIL | ALFFHMYSHN | SSLSKKAQPER | MHVLYLS | SMRSGSS | FFVQ | 60 |
| QY | 61 | LFQGH | PDVFLMEPRAM | VMWTFKOST | AMLMHAYRDL | IRAVFL | CDMSV | PDAYME | PPRRQ 120 |
| DB | 61 | LFQGH | PDVFLMEPRAM | VMWTFKOST | AMLMHAYRDL | IRAVFL | CDMSV | PDAYME | PPRRQ 120 |
| QY | 121 | SSLFQ | MENSRALCSAP | ACDIIPODEI | IPRAHCR | LLCSQGF | FEVEKAC | RSYSHV | LKEVR 180 |
| DB | 121 | SSLFQ | MENSRALCSAP | ACDIIPODEI | IPRAHCR | LLCSQGF | FEVEKAC | RSYSHV | LKEVR 180 |
| QY | 181 | FFNLQ | SLYPLKDP | SLNLHIVL | VRPRAV | FRSRERT | KGDMIS | RIVMQ | HEQKLKED 240 |

Db 181 FENLOSXPPLKDPSPNLNLIHVDPRAVFRSRERITKGDLMIDSRIVMGQHEQKLKED 240
QY 241 QPYVMQVICOSOLEITYKTIOSLPKALOERYLIVREDLARAPYAQTSRMWTEFVGLPELP 300
Db 241 QPYVMQVICOSOLEITYKTIOSLPKALOERYLIVREDLARAPYAQTSRMWTEFVGLPELP 300
QY 301 HLOTWVHNITRGKGMGDHAFHTNARDALNVSQAWRWSLPEYKVSRLQKACGDAMNLLGYR 360
Db 301 HLOTWVHNITRGKGMGDHAFHTNARDALNVSQAWRWSLPEYKVSRLQKACGDAMNLLGYR 360
QY 361 HVRSEQORNLDDLSTWVPEQIH 386
Db 361 HVRSEQORNLDDLSTWVPEQIH 386

Search completed: January 11, 2003, 01:32:03
Job time : 17 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:30:21 Search time 13 Seconds
(without alignments)
576.062 Million cell updates/sec

Title: US-09-816-825-2
Perfect score: 386
Sequence: 1 MLPPKMKLLFLVSQMAIL.....EQRNLLDLSTWTVEQIH 386

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 118974 segs, 19401057 residues

Word size: 15

Total number of hits satisfying chosen parameters: 9

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database: Published Applications-AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-----------------|
| 1 | 386 | 100.0 | 386 | 9 | US-10-007-262-1 |
| 2 | 386 | 100.0 | 386 | 10 | US-09-816-825-2 |
| 3 | 60 | 15.5 | 171 | 10 | US-09-927-602-8 |
| 4 | 19 | 4.9 | 169 | 10 | US-09-927-602-6 |
| 5 | 19 | 4.9 | 169 | 10 | US-09-927-602-7 |
| 6 | 19 | 4.9 | 390 | 10 | US-09-927-602-4 |
| 7 | 19 | 4.9 | 395 | 10 | US-09-927-602-2 |
| 8 | 19 | 4.9 | 395 | 10 | US-09-927-602-3 |
| 9 | 19 | 4.9 | 418 | 10 | US-09-927-602-5 |

ALIGNMENTS

RESULT 1
US-10-007-262-1
Sequence 1, Application US/10007262
Patent No. US20020164748A1
GENERAL INFORMATION:

```

; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSTL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-007-262-1

```

Query Match 100.0%; Score 386; DB 9; Length 386;
Best Local Similarity 100.0%; Fred. No. 0;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MLPPKMKLLFLVSQMAILALFEHMYSHNSSLKKAQERHVLVLSWRGSSFFVG 60
DB 1 MLPPKMKLLFLVSQMAILALFEHMYSHNSSLKKAQERHVLVLSWRGSSFFVG 60
QY 61 LFGQHPDVEFLMEPAHVMWTFKOSTAMMLHNAVRDLIRAVFLCDMSVFDAYMEGPFRQ 120
DB 61 LFGQHPDVEFLMEPAHVMWTFKOSTAMMLHNAVRDLIRAVFLCDMSVFDAYMEGPFRQ 120
QY 121 SFLFQWNSALCSAPACDIPDEIIPRAHCHLLCSQPFVEVERACSSYSHVLEKVR 180
DB 121 SFLFQWNSALCSAPACDIPDEIIPRAHCHLLCSQPFVEVERACSSYSHVLEKVR 180
QY 181 FENQSLYPLKPSLNLAHVLPDPRAVFRSRETKGDMIDSRIVMGHOFKTKED 240
DB 181 FENQSLYPLKPSLNLAHVLPDPRAVFRSRETKGDMIDSRIVMGHOFKTKED 240
QY 241 QPYVWCVTOSQLEIYKTIQSLPKALOERYLLVRREDLARAFAVQTSMTYEVGIEFLP 300
DB 241 QPYVWCVTOSQLEIYKTIQSLPKALOERYLLVRREDLARAFAVQTSMTYEVGIEFLP 300
QY 301 HLOTWVNIITRGKMGDPHETNARDALNYSQAMRMSLPYEXKSRLOKAGDAMNLIYR 360
DB 301 HLOTWVNIITRGKMGDPHETNARDALNYSQAMRMSLPYEXKSRLOKAGDAMNLIYR 360
QY 361 HVRSQEQRNLLDLSTWTVEQIH 386
DB 361 HVRSQEQRNLLDLSTWTVEQIH 386

```

RESULT 2
US-09-816-825-2
Sequence 2, Application US/09816825
Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSTL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816,825
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045,284
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 386
TYPE: PRT
ORGANISM: Homo sapiens
US-09-816-825-2

Query Match 100.0%; Score 386; DB 10; Length 386;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPRKMLLFLVSQMAILLFFHMSHNTSSLSMKAQPRMHVLVSSWRGSSSFVGQ 60
DB 1 MLPRKMLLFLVSQMAILLFFHMSHNTSSLSMKAQPRMHVLVSSWRGSSSFVGQ 60
QY 61 LFGQPDVFLYMEPRAMHWMFFKOSTAMMLHMAVRDLIRAVFLCDMSVEDAPYMPGRRO 120
DB 61 LFGQPDVFLYMEPRAMHWMFFKOSTAMMLHMAVRDLIRAVFLCDMSVEDAPYMPGRRO 120
QY 121 SLEFGWNSRALCSAPACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVR 180
DB 121 SLEFGWNSRALCSAPACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVR 180
QY 181 FENQSLVPLKDPSLNHIYHLVDRPRAVRSRRTKGDMLDSRIYMGHBEKLEKED 240
DB 181 FENQSLVPLKDPSLNHIYHLVDRPRAVRSRRTKGDMLDSRIYMGHBEKLEKED 240
QY 241 QPYVYMOYICQSOLEIKTTIQSLPRALQERYLLVRYEDLARAPYAQTSRMVEFYGLEFLP 300
DB 241 QPYVYMOYICQSOLEIKTTIQSLPRALQERYLLVRYEDLARAPYAQTSRMVEFYGLEFLP 300
QY 301 HLOQWVHNTKRGKMGDAFTNARDALNYSQAMRNSLPYKRSRLKACGDANLLGYR 360
DB 301 HLOQWVHNTKRGKMGDAFTNARDALNYSQAMRNSLPYKRSRLKACGDANLLGYR 360
QY 361 HVRSQEQRNLLDLSTWTWPEQIH 386
DB 361 HVRSQEQRNLLDLSTWTWPEQIH 386

RESULT 3
US-09-927-602-8
Sequence 8, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 171
TYPE: PR
ORGANISM: Homo Sapien
US-09-927-602-8

Query Match 15.5%; Score 60; DB 10; Length 171;
Best Local Similarity 100.0%; Pred. No. 4,4e-50;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 KACRSYSHVYLKEVRFFNLQSLPLKDPSLNHIYHLVDRPRAVRSRRTKGDMLDS 225
DB 53 KACRSYSHVYLKEVRFFNLQSLPLKDPSLNHIYHLVDRPRAVRSRRTKGDMLDS 112
RESULT 4
US-09-927-602-6
Sequence 6, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 390
TYPE: PR
ORGANISM: Homo Sapien

FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 169
TYPE: PR
ORGANISM: Homo Sapien
US-09-927-602-6

Query Match 4.9%; Score 19; DB 10; Length 169;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVSSWRGSSSFVGOLF 62
DB 10 HVLVSSWRGSSSFVGOLF 28

RESULT 5
US-09-927-602-7
Sequence 7, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 169
TYPE: PR
ORGANISM: Homo Sapien
US-09-927-602-7

Query Match 4.9%; Score 19; DB 10; Length 169;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ACRSYSHVYLKEVRFFNLQ 185
DB 51 ACRSYSHVYLKEVRFFNLQ 69

RESULT 6
US-09-927-602-4
Sequence 4, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 390
TYPE: PR
ORGANISM: Homo Sapien

Mon Jan 13 09:14:03 2003

us-09-816-825-2.0115.rapb

Page 3

US-09-927-602-4

Query Match 4.9%; Score 19; DB 10; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 ACRSYSHVLEKVEFFNLQ 185
|||||
Db 165 ACRSYSHVLEKVEFFNLQ 183

RESULT 7

US-09-927-602-2
; Sequence 2, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-2

Query Match 4.9%; Score 19; DB 10; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVLSMRSGSFFVGOLF 62
|||||
Db 42 HVLVLSMRSGSFFVGOLF 60

RESULT 8

US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
; NAME/KEY: VARIANT
; LOCATION: (1)...(395)
; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Query Match 4.9%; Score 19; DB 10; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVLSMRSGSFFVGOLF 62
|||||
Db 42 HVLVLSMRSGSFFVGOLF 60

RESULT 9

US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

Query Match 4.9%; Score 19; DB 10; Length 418;
Best Local Similarity 100.0%; Pred. No. 1.8e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HVLVLSMRSGSFFVGOLF 62
|||||
Db 66 HVLVLSMRSGSFFVGOLF 84

Search completed: January 11, 2003, 01:32:30
Job time : 14 secs

This Page Blank (uspto)

QY 130 RALCSAPACDIIIPODEIIPRACHRLCSQOPFEVEKACRSYSHVYLKEVFNLSLY- 188
 DB 1349 -----CTLLPQKDF-----ENLMLKIDKAMONDKILAIISVSELSLWQ 1390
 QY 189 -----PLKDPSPINLHVLPDPRVFRSRRRTKGLM-----IDSRIVMQQ 231
 DB 1391 EIEMLKFRLESTDAQMGIRLKGISFQVFRQHFLLTKKDLKALVENIDMTSLIL-- 1448
 QY 232 HEQKLRKEDPPYVMQYICQSQLEIKTIOSLPKALQERLYLVRYEDLARAPVAGTSRMV 291
 DB 1449 -----EYCSGFQDCCAVQLF----- 1465
 QY 292 EFVGLFEFLPLQTVHNITRKGMDHAFHTNAR 325
 DB 1466 -----IETLLHNTNAGCGGASMSAKR 1489

RESULT 15
 ID TPSA_CAEEL STANDARD; PRT; 380 AA.
 AC 077081; Q9NEM9;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein-tyrosine sulfotransferase A (EC 2.8.2.20) (Tyrosylprotein
 sulfotransferase-A) (TPST-A).
 GN Y11B2A.15.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=98406128; Pubmed=9733778;
 RA Ouyang Y.-B., Moore K.L.;
 RT "Molecular cloning and expression of human and mouse tyrosylprotein
 sulfotransferase-2 and a tyrosylprotein sulfotransferase homologue in
 Caenorhabditis elegans.";
 RL J. Biol. Chem. 273:24770-24774(1998).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Sulston J.E.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN REVISTONS.
 RA Durbin R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
 CC ACIDIC MOTIFS OF POLYPEPTIDES.
 CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + protein tyrosine =
 CC adenosine 3',5'-bisphosphate + protein tyrosine-O-sulfate.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By
 CC similarity).
 CC -!- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF049709; AAC36062.1; -
 DR EMBL; AL132904; CAC35844.1; -
 DR WormPep; Y11B2A.15; CE26632.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 KW Transferase; Transmembrane; Glycoprotein; Signal-anchor.
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 7 27 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 28 380 LUMENAL, CATALYTIC (POTENTIAL).
 FT CAROHRD 66 66 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 380 AA; 43313 MW; FF709BF0F1EDC95 CRC64;

Query Match 4.3%; Score 87.5; DB 1; Length 380;
 Best Local Similarity 18.1%; Pred. No. 3.8;
 Matches 79; Conservative 52; Mismatches 117; Indels 189; Gaps 19;

QY 5 KMKLLFLVSMALALAEFHM-----YS-----HNISLS 35
 DB 5 RELILVFLV-----VFILFYPTARTADDPYSSNHRKFNCAADGDSESLPFHQLSVR 60
 QY 36 MKAQPERKHVVLSSW--RSGSFFVGOLFQGHPDV-----FYLMERAW-----HV 78
 DB 61 SDDGYNRTSPFFIFIGVPRSGTTLIRAMIDAPREYRCCEETRVIPRIILNLSQMKKSEKE 120
 QY 79 WMTEKQS--TAMMLHMAVEDLIRAVFLCDMSVEDAYMEGPPRQSSLFQWENSRAICGAP 136
 DB 121 WNRLOQAGVTGEVINNAISSFI-----MEIMVGHGDRAPR----- 155
 QY 137 ADDIIPODEIIPRACHRLCSQOPFEVEKACRSYSHVYLKEVFNLSLYPLKDPSTL 196
 DB 156 -----LCNADPFTM-----KSAYILKE-----LFP----- 175
 QY 197 NLHIVLVPDPRVFRSRRRTKGLMDTSRIV-----MGQHECKLRKEDPPYVMQYIC 250
 DB 176 NAKYLLMIRDSGRATVNS-----IISRKVITIGFDLNDPQCMTRWNAIQLMVDQC 226
 QY 251 QSQLEIKTIQSLPKALQERLYLVRYEDLARAPVAGTSRMVFEVLEFLPHQTVHNIT 310
 DB 227 ESVGE-----KNCLEKYYEQLVLHPEAKMRITETLDP----- 260
 QY 311 RKGMDHAFHTNARDALNVSAQWMSLPEKVSRLQACGDMNLGVRHVSDEQGRN 370
 DB 261 -----WD-----DKVLMHEQLIGDISLSNVR-SSQGVVVKP 291
 QY 371 LLIDLISLV--TVPEQI 385
 DB 292 VNIDALIKKVGITPEDEV 308

Search completed: January 11, 2003, 01:19:27
 Job time : 39 secs

Best Local Similarity 20.7%; Pred. No. 24;
Matches 71; Conservative 38; Mismatches 108; Indels 126; Gaps 18;

```

QY 23 FFMVSHNSSLKAKOPERMHVLVSSWSSGSSFGQLFGQHPDY-----68
Db 367 FARVINH-SSMCATIMEPOL-----KEFMGNWKGKMSVLAARSSVRAVCVN 414
QY 69 -----FLMEPAHVMWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVDAWMEPG 116
Db 415 KFTNGKLPYSTTLNLNETW--WNIWENSTAMF-----FDTAEVD 453
QY 117 -PROSSLFOWENS--RALCSAPACDIIPODEIIPRAHCRLLC--SQQPEVEVER---166
Db 454 VPEKLDLHFMGEGAGLVAAHITSRYVGTVP---LADRENMALLCMDSQKLIHMRFRMR 510
QY 167 -ACRSYSHVYLKVRFPNLSLYPLKDPSLNLIHVLVADPRVPRRSRTKGDMLIDS 225
Db 511 GAWGAMHCVCISRE--FLULYVEARLKSCL-----IAARRR-----545
QY 226 RIVMGQHEOR-----LKKEDQPYVMQVICOSELEIKTIQSLPKALOERYLLVRE 277
Db 546 ---GQHKELKLEAVEVLGKSSDALFRAMTYLCLMARLEPMFSESGL-----RFFLLRGR 595
QY 278 DLARAPVQTSRMEFEGLEFLPHLQTW---VANIT--RGKGM 315
Db 596 NNLXGLTNYTEGKRAVTGVNL---WSNVHEVSTKRHKM 633

RESULT 13
F3ST_FLABI
ID F3ST_FLABI STANDARD: PRT: 312 AA.
AC P52835;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Flavonol 3-sulfotransferase (EC 2.8.2.-) (F3-ST).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; easterids II; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Flaveria.
OX NCBI_TaxID=4224;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95083751; PubMed=7991681;
RA Manoveranich S., Varin L., Gullick P., Ibrahim R.;
RT "Cloning and regulation of flavonol 3-sulfotransferase in cell-
RT suspension cultures of Flaveria bidentis.";
RL Plant Physiol. 106:485-491(1994).
CC -I- FUNCTION: CATALYZES THE SULFATE CONJUGATION OF QUERCETIN.
CC RHAMNETIN AND ISORHAMNETIN BUT NOT KAEMPEROL. O-SULFATION OF
CC POSITION 3 OF FLAVONOL. MAY PLAY A ROLE IN AUXIN TRANSPORT.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- TISSUE SPECIFICITY: HIGHEST IN SHOOT TIPS AND LOWEST IN MATURE
CC LEAVES AND ROOTS.
CC -I- SIMILARITY: BELONGS TO THE PLANT SULFOTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: U10275; AAA61638.1; -
DR HSSP: P50224; ICDM;
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransferase; 1.
DR ProDom: PD001218; Sulfotransferase; 1.
KW Transferase.
KW BINDING 138 153 PAPS-BINDING SITE (BY SIMILARITY).
SQ SEQUENCE 312 AA; 36457 MW; 29D17458CB663FC2 CRC64;

```

Query Match 4.3%; Score 88; DB 1; Length 312;

Best Local Similarity 21.7%; Pred. No. 2.7;
Matches 54; Conservative 41; Mismatches 86; Indels 68; Gaps 13;

```

QY 60 QLEGGHP-DVEFLMEPAHVMWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVDAWMEPGPR 118
Db 43 QTEKAPNPNVFLASYP-----KSGTTLKALAFAITREKF-----78
QY 119 ROSSLFOWENSRALCSAPACDIIPODEIIPRAHCRLLCSQQPEVEVERKACRSYSHVYLKE 178
Db 79 -----DSS-----TSPLLTTPHD-CIP-----LL--EKDLEKIOEORNSLYTPIS- 117
QY 179 VRFENIQLSYPLKDPSLNLIHVLVADPRVPRRSRTKGDMLIDSRIVMQHEQKILK 238
Db 118 -THFHKSLPESAR--TSCKIYIYIRMKDYIVSYHPLRQIV-----KLSV 162
QY 239 EDQPY-YVMQVICO--SOLEIYKTIQSLPKALOER---YLLVREYDLARAPVQTSRMY 291
Db 163 EEAPEEAVDERFCQGISCGPYWEHTLGYMKASLEKPEIFLEKEDMKKDPVPSVKRLA 222
QY 292 EFVGLLEFLP 300
Db 223 DFTGHPFTP 231

RESULT 14
Y166_HUMAN
ID Y166_HUMAN STANDARD: PRT: 2209 AA.
AC P50748;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein KIAA0166.
GN KIAA0166.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: D79988; BA011483.1; -
DR Genew: HGNC:17255; KNTC1.
KW Hypothetical protein.
SQ SEQUENCE 2209 AA; 250746 MW; 51AC948733CBFDAC CRC64;

Query Match 4.3%; Score 88; DB 1; Length 2209;
Best Local Similarity 16.2%; Pred. No. 33;
Matches 54; Conservative 46; Mismatches 98; Indels 134; Gaps 10;

```

```

QY 29 HNSSLKAKOPERMHVLVSSWSSGSSFGQLFG--QHPDVFL-----M 72
Db 1253 NISIALLDMLQDS-----SQWELALRFVVGSGFTCLQHSVSNMNAITSEKILGETTL 1305
QY 73 EPANVHMWTFKOSTAMMLHMAVRDLIRAVFLC---DMSVPAYMEPGRROSSLFWENS 129
Db 1306 VKSRHVVELKEKAVIFIRENATTLILHRVFNCRCLVDLDLALGY-----1348

```

```
QY      70 YLM--EPAW---HWMTFKO-----STAMMLHNAVDLIRAVFLCDMSVFDAIMEBG   116  
| | : ||||         |||           | : | : | : | : | : | : | :  
  
Db      198 VEEACARSHSYSEVILKEVEFFNLGSLYPLKDPISNLIHLVLVPDRAFRSRERTKGIDL    222  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB      134 VLPRSS-----LL-----ALNCNKVVLYRNIXDV-----          156  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
  
QY      223 IDSRIWGQHOKKKREDDPIYYMVVICQSOLEIKTIQSIIP-----RALQR--     270  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db      157 IVSFHFGRRTTKLPLEDPAPF-----EAEDFEFHAGLSGPCPYWDHLGLGYWKASLERPE  210  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
QY      271 -YLIVRYEDLARAPAOTSRMYEFVGIEF       298  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db      211 VIIFKIYDKRKDPTSNVKRLAEFIIGTF        239  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
  
RESULT 11  
MPA_RAT MPPA_RAT STANDARD; PRT; 524 AA.  
P20063;  
DT 01-FEB-1991 (Rel. 17, Created)  
DI 01-FEB-1991 (Rel. 17, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DT Mitochondrial processing peptidase alpha subunit,  
precursor (EC 3.4.24.64) (Alpha-MPP) (P=55).  
GN PMPCA OR MPPA..  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI_TaxID=10116;  
[1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=91045920; PubMed=2236012;  
RA Kleiber J., Kalousek F., Swarcop M., Rosenberg L.E.;  
RT "The general mitochondrial matrix processing protease from rat liver:  
structural characterization of the catalytic subunit." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 87:7978-7982(1990).  
CC -I- FUNCTION: THE MITOCHONDRIAL PROCESSING PROTEASE (MPP-I) CLEAVES PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS. MOST MPP-I CLEAVAGE SITES FOLLOW AN ARGinine AT POSITION -2.  
CC -I- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from precursor proteins imported into the mitochondrion, typically with Arg in position p2.  
CC CC SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS.  
CC -I- SUBCELLULAR LOCATION: Mitochondrial matrix.  
CC -I- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE ZINC-BINDING SITE.  
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M6.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/or send an email to license@isb.sib.ch).  
CC DR EMBL: M57728; AAA41632.1; ..  
DR PTR: A36205; A36205..  
DR MEROPS: M6_971;  
DR InterPro: IPR001431; Peptidase_M6.  
DR Pfam: PF00675; Peptidase_M6; 1.  
DR ProSITE: PS00143; INSULINASE; 1.  
KM Hydrolyse; Metalloprotease; Mitochondrion; Transit peptide.  
FT TRANSIT 1 32 MITOCHONDRION.  
FT CHAIN 33 524 MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA SUBUNIT.  
SQ SEQUENCE 524 AA: 56607 MW; 8BF0BFCGFDF09DB2 CRC64;
```

```

Db      269  YLLGQPMWCAAGAAWMLTQAHNSRGSSRW-----RE-----TCOMSA---LRP 311
OY      117  PRROSLTQWENSRALCSAPACDITIPQDEIFRAICRLC-----SQQPEFVE 165
Db      312  PRFOSSH-IYGGARL-----LLEDEDFIPFVLANMMGGGSGFSAGGKGKMSF--- 360
OY      166  KACRSYSHVVLKEVEFENIQLSLYPLKDPSTLHTIVLDRPRAVRSRETRKGLMIDS 225
Db      226  RIVMGHQKRLKKEQPPYVQVITQSQDEIYIKTIQSLPKALQERYLLVRYEDLARAYV 285
OY      411  FILMRVTDVLELEFAKTQLMSLMKN-----LESRP-----VIEDVRQYLA 454
Db      286  QTSRYEYFVGLLEFLRH-LQTWVHTTRCKGMGDAHFHTNARDALNVSQAMWLSLEYKVS 344
OY      455  THSRK-----LPHELTCLIRNV-----KPEDIKRVASKMLGRP----- 488
Db      345  RLQKACGDAMNLLGYRWVSEQEOEN 370
OY      489  -AVALGDLTDLPTEYHTQALSRD 513

RESULT 12
RRPO_PVMR STANDARD: PRT: 1968 AA.
AC P17965; Q89548;
DT 01-NOV-1990 (Rel. 16, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA replication protein (147 kDa, pretein) (ORF 1) [Contains: RNA-
directed RNA polymerase (EC 2.7.7.48); Probable helicase].
OS Potato virus M (strain Russian) (PVM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
OX NCBI_Taxid=12168;
RN [1]
RRP SEQUENCE FROM N.A.
RX MEDLINE=91116326; PubMed=1990070;
RA Zaytsev S.K., Kanyuka K.V., Leay K.E.;
RT "The genome organization of potato virus M RNA.";
RL J. Gen. Virol. 72:9-14(1991).
CC -I- FUNCTION: RNA-replication. The central part of this protein
possibly functions as an ATP-binding helicase.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
{RNA}(N).
CC -I- SIMILARITY: CONTAINS 1 OTU DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D14449; BAAC0339.1; -.
DR PIR: S21601; S21601.
DR PIR: PNO093; PNO093.
DR MEROPS: C23.001; -.
DR InterPro: IPR003323; OTU.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR000606; Viral_helicase1.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF02388; OTU; 1.
DR PROSITE: PS50802; OTU; 1.
KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW Transferrase.
KW DOMAIN 883 991 OTU.
FT NP_BIND 1166 1173 ATP (POTENTIAL).
SQ SEQUENCE 1968 AA; 223384 MW; 6F15A7951AD96AAC CRC64;

Query Match 4.4%; Score 89; DB 1; Length 1968;

```

Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A., Viari A., Wambuit R., Wedler H., Weitzengger T., Winters P., Wipst A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zurchin A.;
 "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*,"
 Nature 390:249-256(1997).
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE OF 1-1763 FROM N.A.
 RC STRAIN-168 / PBI424;
 RA Tognoni A., Grandi G.;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
 CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
 CC SECONDARY METABOLISM.
 CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES
 CC (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z99113; CAB13603.1; -
 DR EMBL: Z35133; CA84505.1; -
 DR Subtilist: BG10931; PKSM.
 DR InterPro: IPR002198; ADH_short.
 DR InterPro: IPR00794; Ketoacyl-synt.
 DR InterPro: IPR001601; Methyltransf.
 DR InterPro: IPR003880; Ppantne_attach.
 DR InterPro: IPR000051; SAM_bind.
 DR Pfam: PF00106; adh_short; 1.
 DR Pfam: PF00109; ketoacyl-synt; 3.
 DR Pfam: PF00550; pp-binding; 4.
 DR Pfam: PF02801; ketoacyl-synt_C; 3.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.
 DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE: PS50075; ACP_DOMAIN; 4.
 KM Transferrase; Acyltransferase; Multifunctional enzyme; Repeat; Complete proteome.
 KW Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
 FT DOMAIN 295 364 ACYL CARRIER (ACP) 1.
 FT DOMAIN 396 834 BETA-KETOACYL SYNTHASE 1.
 FT DOMAIN 2190 2258 ACYL CARRIER (ACP) 2.
 FT DOMAIN 2322 2737 BETA-KETOACYL SYNTHASE 2.
 FT DOMAIN 3532 3947 BETA-KETOACYL SYNTHASE 3.
 FT DOMAIN 3410 3483 ACYL CARRIER (ACP) 3.
 FT DOMAIN 4140 4209 ACYL CARRIER (ACP) 4.
 FT BINDING 327 327 PHOSPHOPANTETHEINE (POTENTIAL).
 FT BINDING 2222 2222 PHOSPHOPANTETHEINE (POTENTIAL).
 FT ACT_SITE 2476 2476 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT BINDING 3446 3446 PHOSPHOPANTETHEINE (POTENTIAL).
 FT ACT_SITE 3690 3690 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT BINDING 4172 4172 PHOSPHOPANTETHEINE (POTENTIAL).
 DR CONFLICT 103 103 E -> V (IN REF. 2).
 DR CONFLICT 276 276 T -> S (IN REF. 2).
 FT CONFLICT 289 289
 SQ SEQUENCE 4273 AA; 477459 MW; 3BBFCF1A250AEB5A CRC64;
 Query Match 4.4%; Score 89.5; DB 1; Length 4273;
 Best Local Similarity 22.9%; Pred. No. 58;
 Matches 48; Conservative 34; Mismatches 87; Indels 41; Gaps 10;

OY 276 YEDLAPAPAOOT-SRMYE-----FYGLEFPHLOTWNINIT--RKGMDGHAFHTN 323
 DB 143 SGEIVTGMKARGQIYEAKEGAVIDLAVGQALRHSDAFPHPTLIDSGTGS----- 196
 OY 324 ARDALNVQAMRWNSLPEYKVS---RLQKAC 350
 DB 197 --SCLISDQTWLPLIYESFSSASERLQKCC 224

RESULT 10
 FAST_FLACH STANDARD: PRT; 320 AA.
 ID P52837;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Flavonol 4'-sulfoltransferase (EC 2.8.2.-) (F4-ST).
 OS *Flavaria chloraeifolia*.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
 OC Helianthaceae; Flaveria.
 OX NCBI_TaxID=4228;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92159034; PubMed=1741382;
 RA Varin L., Deluca V., Ibrahim R.K., Brisson N.;
 RT "Molecular characterization of two plant flavonol sulfoltransferases,"
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1286-1290(1992).
 RN [2]
 RP PAPS-BINDING SITE.
 RX MEDLINE=95279378; PubMed=7759495;
 RA Varin L., Marsolais F., Brisson N.;
 RT "Chimeric flavonol sulfoltransferases define a domain responsible for
 RT substrate and position specificities,"
 RL J. Biol. Chem. 270:12498-12502(1995).
 CC -1- FUNCTION: TRANSFERS SULFATE GROUP INTO FLAVONOLIN POSITION 4'.
 CC MAY PLAY A ROLE IN AUXIN TRANSPORT.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: HIGHEST IN SHOOT TIPS AND LOWEST IN MATURE
 CC LEAVES AND ROOTS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PLANT SULFOTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M84136; AA83343.1; -
 DR HSBP; P50224; ICDM.
 DR InterPro: IPR000863; Sulfoltransferase.
 DR Pfam: PF00685; Sulfoltransfer; 1.
 DR PRODOM: PD001218; Sulfoltransferase; 1.
 KW Transferrase.
 FT BINDING 148 163 PAPS-BINDING SITE.
 SQ SEQUENCE 320 AA; 37255 MW; 6B263659F6CCBC0 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 320;
 Best Local Similarity 20.8%; Pred. No. 2.3;
 Matches 56; Conservative 30; Mismatches 83; Indels 100; Gaps 12;

DR PROSITE: P500410; DYNAMIN: 1.
 KW Interferon induction; GTP-binding; Multigene family.
 FT NE_BIND 75 82 GTP (POTENTIAL).
 FT NE_BIND 176 180 GTP (POTENTIAL).
 FT NE_BIND 245 248 GTP (POTENTIAL).
 SQ SEQUENCE 659 AA; 74951 MW; F55D632838C865B3 CRC64;
 Query Match 4.5%; Score 91; DB 1; Length 659;
 Best Local Similarity 20.6%; Pred. No. 3.9;
 Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;
 QY 62 FGQHPDVFYLMERAMHWMTFKOSTAMLMHAYRDLI---RAVFLCDMSVDPAYME--- 114
 DB 301 FKEHPERALLEGG-----KAIVPOLARLITWELISHCKSLPLENOIKESHSTSE 353
 QY 115 -----PGRPROSLFQWENSRALCSAPACDIIPODEIPRACRLCS-QCPF--- 161
 DB 354 ELQYGADIPEDENKFLIEKINAF-NODITAVEGEIVREKRCRLFTKLRKEFFLM 412
 QY 162 -EYVERKCRYSHVLRKEAFVNLQ-----SLYLRKPSLNLI 200
 DB 413 SELEERFQSGSALYKEVTFEMQYRGRELPGFVNYKTFENTIRQIKLEPPAME- 470
 QY 201 VHLVRD-PRAVFR-----SRETKGDLMDISRIYMGHEOKLKREDQPYVMQV 248
 DB 471 LHWTEIVRAFLTVSEKNSEFFNLRTKSKLEDRL---EQITAEAKALHQMED 527
 QY 249 ICOSOLEIYKTQSIPKALOERYLVRYEDL-----ARAPVAQTSRMVEFVGL 296
 DB 528 IIVCODIYR-----KALOK---VREEAEEERKHGKRSASQSPNQTSSMD----- 572
 QY 297 EFLPHLOTW---VEN 308
 DB 573 EIFQHLNAYRQEAHN 587
 RESULT 8
 C7D8-SOYBN STANDARD; PRT; 504 AA.
 ID C7D8-SOYBN STANDARD; PRT; 504 AA.
 AC 081974;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE Cytochrome P450 71D8 (EC 1.14.-.-) (P450 CPT).
 OS Glycine max (Soybean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 OX NCBI_Taxid=3847;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Harosoy 63;
 RX MEDLINE=98311068; PubMed=9648734;
 RA Schopfer C.R., Ebel J.;
 RT "Identification of elicitor-induced cytochrome P450s of soybean
 (Glycine max L.) using differential display of mRNA."
 RL Mol. Gen. Genet. 258:315-322(1998).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC CC
 CC EMBL: Y10493; CAA7517.1; -
 DR HSSP: P14779; 10PZ.
 DR Interfero: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: P500086; CYTOCHROME_P450; 1.

KW Oxidoreductase; Monooxygenase; Heme.
 FT BINDING 444 444 HEME (BY SIMILARITY).
 SQ SEQUENCE 504 AA; 57579 MW; 8654FAEEC2FF2A6F CRC64;
 Query Match 4.4%; Score 89.5; DB 1; Length 504;
 Best Local Similarity 19.9%; Pred. No. 3.7;
 Matches 70; Conservative 45; Mismatches 108; Indels 129; Gaps 16;
 QY 68 VFYIMEPAMHWMTFKOSTAMLMHAYRDLIRAVFLCDMSVDPAYMEGRPROSLFQME 127
 DB 14 VELLH--WLV-KTYKQKSSHL-----PGPMR----- 39
 QY 128 NSRALCSAPACDIIPODEIPRACRLCSQDPFEVVERKCRYS--HYLVKEVFNLI 184
 DB 40 -----LP---IIGNLHQLALASLPDGLKLVKRGPLMHLDGLSTLV 83
 QY 185 QSLVPLKDPSSLNLHVLVPRPRAVFRSRETKGDLMDISRIYMGHEOKLKREDQPY 244
 DB 84 SS--FKMAEMKKTIDVHVRQ-----LLAPQMYGARDIAFAPRGDYWR 129
 QY 245 VMOVTCOSOLEIYKTQSIPKALQ--RYLVRYEDLARAPVAQTSRMVEFVGL----- 296
 DB 130 QIRKICTLELLSAKRVQSFSHRQDENKLLQSHSSAGSPIDLSGLFSLGTVSRAA 189
 QY 297 -----EFL-----PHLOTWNYNIRKGMGDHFTNA----- 324
 DB 190 FKGENDDDQEFNSLVKAKALITMTGFEVDMPSLKP-LHLTRQAKAEHV-HQRAQKIL 247
 QY 325 RDALNVSQAWMSLEPKYKSLQKACGDAMNLGYRHSQEQDNLDDL 376
 DB 248 EDLRKHM-----EKRTVRKEGNG-----SEAEQDLVDVLL 279
 RESULT 9
 PRSM-BACSU STANDARD; PRT; 4273 AA.
 ID PRSM-BACSU STANDARD; PRT; 4273 AA.
 AC P40872; O31781;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2003 (Rel. 41, Last annotation update)
 DE Putative polyketide synthase pksM.
 OS PksM OR pksY.
 GN Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_Taxid=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloul G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriess R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Ertlan K.D., Erington J., Faber C., Ferrari E., Foudget D.,
 RA Fritz C., Fujita M., Fujita Y., Fume S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Goldightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kashara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Larois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nockack M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudea B., Park S.H.,
 RA Parro V., Pohl T.M., Portetle D., Porciliak S., Prescott A.M.,
 RA Preece E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rochna E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scofield F.,
 RA Sekiguchi Y., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,

Best Local Similarity: 22.2%, Pred. No.: 53;
Matches: 88; Conservative: 66; Mismatches: 128; Indels: 115; Gaps: 24

```
OY      2 LLPKMKLLPLVSOAMTALTFPHNTSHNISLSMKAQPERNNVLVLSWRSQSSFGQL   61  
          ||::|||::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     1732 ILISQQNFIATQSQA-----FLDQHGNLT-----PEEQML-----OQTGL    1792  
  
OY      62 FGDHDVEFLMRPWMTWTGKSTAMIMHM-AVRDLRAVLICMSVDAYMEPGRRO   120  
          ::|::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     1793 KEQYS-----TSLASDAELKKOVTLDELQK-FLDHKEESMLE---RSE   1835  
  
OY      121 SLFWENSRFALCSAPACDIIPDEITPRANCRLSCSQPF--EWEKACSRYSH----   173  
          |::|::|::|::|::|::|::|::|::|::|::|::|:  
Db     1836 KEL---ENNHKGGSSP-----ETLP---SLLRQGSSFEDVI-----SKRGDLR   1873  
  
OY      174 -VLLEVEFPFNOSLYPLDKDSPLMHLYHYIVRDRAWRSERFKGMIDMR-----   226  
          :|::|::|::|::|::|::|::|::|::|::|::|:  
Db     1874 FYTIIGQXVLDIENSFFCKEKPS---ELGNLVKD---KLDAERYTAALHSKCPRLGS   1925  
  
OY      227 ---IWMGHNEOKLKREDPPYYVVGYVICOSQLE-IYKTIOSLPKALOERNLILVREYDLAR   281  
          ::||::|::|::|::|::|::|::|::|::|::|::|:  
Db     1926 HLNMLLGCVHGQRNSADSIQAAMMQ-CEANVEKLSDTYAASPGLVQE-----        1973  
  
OY      282 APVAQTSKNRYEVGVGLELP--HLQTMWHNI--TRKGMDFAFHINARA-LNVSAWRN   336  
          |::|::|::|::|::|::|::|::|::|::|::|:  
Db     1974 --LATTKQDQELAHOVVERVKQAVARDIMEIEBPAPDRHNVOETTDSILSFQSYIS   2031  
  
OY      337 SLPYEKVERLOKACGDAMNLLEYREVSRHQEOBNLL   373  
          ||::|::|::|::|::|::|::|::|::|:  
Db     2032 SLA-ERSSILOKAIAGS-----QSVQSELSEL   2058  
  
RESULT 6  
239F_HUMAN  
ID       239F_HUMAN           STANDARD:         PRT:      294 AA.  
AC       Q15777:  
DT       15-JUL-1998 (Rel. 36, Created)  
DT       15-JUL-1998 (Rel. 36, Last sequence update)  
DT       16-OCT-2001 (Rel. 40, Last annotation update)  
DE       Fetal brain protein 239 (239FB).  
GN       C11ORF8.  
OS       Homo sapiens (Human).  
OC       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC       Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
CX       NCBI_TaxId=9606;  
RN       [1]  
RP       SEQUENCE FROM N.A.  
RX       MEDLINE=95080775.; PubMed=7527372;  
RA       Schwartz F., Neve R., Eisenman R., Gessler M., Bruns G.;  
RT       "A WAGR region gene between Pax-6 and FSB expressed in fetal brain.";  
RL       Hum. Genet. 94:658-664(1994).  
RN       [2]  
RP       SEQUENCE FROM N.A.  
RX       MEDLINE=96115606.; PubMed=8666403;  
RA       Schwartz F., Eisenman R., Knoll J., Gessler M., Bruns G.;  
RT       "cDNA sequence, genomic organization, and evolutionary conservation  
of a novel gene from the WAGR region.";  
RL       Genomics 29:526-532(1995).  
CC       - TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN FETAL BRAIN.  
CC       - SIMILARITY: BELONGS TO THE UZF0046 FAMILY.  
CC       -----  
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC       between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC       The European Bioinformatics Institute. There are no restrictions on its  
CC       use by non-profit institutions as long as its content is in no way  
CC       modified and this statement is not removed. Usage by and for commercial  
CC       entities requires a license agreement (See http://www.isb-sib.ch/annouce/  
CC       or send an email to license@isb-sib.ch).  
CC       -----  
DB     EMBL: U57911; AAC50564.1; -  
DB     Genev: HGNC:1180; C11orf8.  
DB     MW, 600911, -  
DB     InterPro: IPR004843; M-pestrase.
```

```
DR InterPro: IPR004844; S/T_phosphatase.  
DR Pfam: PF00149; Metallophos.1.  
SQ SEQUENCE 294 AA: 33360 MW: 4352BC0DA1BD1F0 CAC64:  
  
Query Match 4.5%; Score 91; DB 1; Length 294;  
Best local similarity 20.2%; Pred. No. 1.4;  
Matches 53; Conservative 30; Mismatches 89; Indels 90; Gaps 9;  
  
QY 139 DIIPDEITFRACHCLLCSQPPEVEVERAKRSYSHYVL-----KEYRFFNLSTL 187  
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 45 DEIPDTPRPAGCHTFVCTSDTHSRDTGIOMPYGDILLTGTGFTGLPSEVKFN---- 100  
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
QY 188 YPLLDPSPSLNIHIVLVRRPAVFRRSREKTKDDLIDSRIYM-GOHEQKLXKE-----D 240  
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 101 -----DWLGNNLPYEKKIYVAGNHLEFPDKFEFMADLVK 132  
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
QY 241 QPIYYMQVVICOSQLTEYKTIGSLPALDERILLYVEDIARAPVAQTSM---YEEVGL 296  
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 133 QDIYFFPSYSKLRPFDFDVOST-----LTNLSYLDSISVTWKGRIRYA 177  
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
QY 297 EPLPHLQTVNHNITRGKGMDHAFTNARDALNVSGAWMSLPYEKVSRLQKACGDAML 356  
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 178 PWTPEENMGFLPPGGSLD-----KNWIIPEGIDLM-THPPP--- 216  
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
QY 357 LGYRVRSSEQEORNLLDLIST 378  
| | | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
Db 217 LGFRTMWPKELQRVCVELLNT 238  
| | | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |  
  
RESULT 7  
MK3_RAT MK3_RAT STANDARD: PROT: 659 AA.  
1C MK3_RAT  
ID P18590:  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE Interferon-induced GTP-binding protein MX3.  
GN MX3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91056598; PubMed=2173790;  
RA Meier E., Kunz G., Haller O., Arthelter H.;  
RT "Activity of rat Mx proteins against a rhadovirus.";  
J. Virol. 64:6263-6269(1990).  
CC -!- FUNCTION: DOES NOT SHOW ACTIVITY AGAINST INFLUENZA VIRUS OR VSV;  
ALTHOUGH IT ONLY DIFFERS FROM MX2 BY 8 POSITIONS.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- INDUCTION: BY INTERFERONS.  
CC -!- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: X52713; CAA36937.1; -.  
CC PIR: S11737; S11737.  
DR InterPro: IPR001401; Dynamn.  
DR InterPro: IPR000375; Dynamln_central.  
DR InterPro: IPR003130; GED.  
DR Pfam: PF00350; dynamin_1.  
DR Pfam: PF01031; dynamin_2; 1.  
DR Pfam: PF02212; GED; 1.  
DR PRINTS: PR00195; DYNAMIN.  
DR SMART: SM00053; DYNC; 1.  
DR SMART: SM00302; GED; 1.
```

RT "Molecular cloning of macrophin, a human homologue of Drosophila
RT kakapo with a close structural similarity to plectin and dystrophin."
RN Biochem. Biophys. Res. Commun. 264:568-574(1999).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE-20026884; PubMed-10559237.
RA Sun Y., Zhang J., Knaef S.K., Auclair D., Chang M.-S., Liu Y.,
RA Sutherland R., Salgia R., Griffin J.D., Ferland L.H., Chen L.B.,
RT "Molecular cloning and characterization of human trabeculin-alpha, a
RT giant protein defining a new family of actin-binding proteins."
RL J. Biol. Chem. 274:33522-33530(1999).
RN [3]
RP SEQUENCE OF 868-2350 FROM N.A.
RP TISSUE-Brain;
RX MEDLINE-20039619; PubMed-10574462;
RX Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345(1999).
RN [4]
RP SEQUENCE OF 1544-5057 FROM N.A.
RP Corby N.;
RN Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 3734-5430 FROM N.A.
RP TISSUE-Brain;
RX MEDLINE-98116662; PubMed-9455484;
RX Seki N., Ohira M., Nagase T., Ishikawa K.-I., Miyajima N.,
RA Nakajima D., Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain".
RL DNA Res. 4:345-349(1997).
RN [6]
RP FUNCTION: F-ACTIN-BINDING PROTEIN WHICH MAY PLAY A ROLE IN CROSS-
RP LINKING ACTIN TO OTHER CYTOSKELETAL PROTEINS. ALSO BINDS TO
RN MICROTUBULES (BY SIMILARITY).
RN [7]
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 37 SPECTRIN REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AB029290; BAA83821.1; -;
DR EMBL: AF141968; AAF06360.1; -;
DR EMBL: AB033077; BAA86565.1; -;
DR EMBL: AL137853; CAC1520.1; -;
DR EMBL: AB007934; BAA32310.1; -;
DR HSSP: Q01082; 1BKR.
DR Genew: HGNC:13664; MACF1.
DR Interpro: IPR001589; Actinbind_actin.
DR Interpro: IPR001715; Calponin-like.
DR Interpro: IPR002048; EF-hand.
DR Interpro: IPR003108; GAS2.
DR Interpro: IPR001452; SH3.
DR Interpro: IPR002017; Spectrin.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF00307; CH; 2.
DR Pfam: PF00435; Spectrin; 36.
DR Pfam: PF02187; GAS2; 1.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00033; CH; 2.
DR SMART: SM00054; Efn; 2.

DR SMART: SM00243; GAS2; 1.
DR SMART: SM00150; Spec; 35.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE: PS00021; CH; 2.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00002; SH3; FALSE_NEG.
KM Actin-binding; Cytoskeleton; Calcium-binding; Repeat; SH3 domain.
FT DOMAIN 1 295
FT 78 181
FT DOMAIN 194 295
FT REPEAT 314 355
FT REPEAT 591 623
FT REPEAT 680 784
FT REPEAT 786 800
FT DOMAIN 871 923
FT REPEAT 1250 1272
FT REPEAT 1287 1342
FT REPEAT 1455 1534
FT REPEAT 1547 1659
FT REPEAT 1815 1891
FT REPEAT 1932 2042
FT REPEAT 2260 2280
FT REPEAT 2372 2395
FT REPEAT 2398 2507
FT REPEAT 2510 2618
FT REPEAT 2621 2728
FT REPEAT 2731 2838
FT REPEAT 2841 2945
FT REPEAT 2967 3024
FT REPEAT 3136 3163
FT REPEAT 3187 3274
FT REPEAT 3277 3383
FT REPEAT 3386 3492
FT REPEAT 3495 3601
FT REPEAT 3604 3619
FT REPEAT 3713 3819
FT REPEAT 3832 3927
FT REPEAT 3982 4043
FT REPEAT 4046 4152
FT REPEAT 4155 4262
FT REPEAT 4265 4371
FT REPEAT 4374 4481
FT REPEAT 4484 4580
FT REPEAT 4593 4700
FT REPEAT 4707 4808
FT REPEAT 4811 4917
FT REPEAT 4920 4985
FT REPEAT 5036 5107
FT CA_BIND 5132 5143
FT CA_BIND 5276 5283
FT DOMAIN 5355 5370
FT CONFLICT 1 72
FT 575 594
FT CONFLICT 1487 1487
FT CONFLICT 1963 1963
FT CONFLICT 2052 2052
FT CONFLICT 2083 2083
FT CONFLICT 2220 2220
FT CONFLICT 2231 2231
FT CONFLICT 2343 2363
FT CONFLICT 2344 2350
FT CONFLICT 2523 2523
FT CONFLICT 4670 4670
FT CONFLICT 4833 4833
SQ SEQUENCE 5430 AA; 620346 MW; 91ADB7F7590B440B CRC64;
Query Match 4.5%; Score 91.5; DB 1; Length 5430;
VAISSSEDEGNIREYVELLS -> GPSAPLKKKAISDLCNM
YCL (IN REF. 1).
A -> T (IN REF. 2 AND 3).
V -> A (IN REF. 1).
E -> D (IN REF. 1).
E -> K (IN REF. 2).
M -> Y (IN REF. 2).
C -> Y (IN REF. 1).
MISSING (IN REF. 2).
SLPSVG -> EYLFKI (IN REF. 3).
O -> R (IN REF. 2 AND 4).
S -> T (IN REF. 2).
MISSING (IN REF. 2).
4 X 4 AA TANDEN REPEATS OF [GS]-S-R-[AR].
MSSDEFTLSRSCSRSRSGSYRSGSLSPGPD
TLPMPLPHEOKRKRSQSDVDPAPRAVAV -> MPPEYLVN
AGIPRGVGSIQPLPRGKQCTASRAVAVI (IN REF.
2).

| | |
|----|--|
| CC | -1- FUNCTION: REQUIRED FOR THE FORMATION OF SULFATED NOD FACTOR. |
| CC | PROPOSED TO TRANSFER ACTIVATED SULFATE (PARS) TO THE FUOSE |
| CC | OF THE NOD FACTOR. |
| CC | -1- SIMILARITY: LIMITED TO NODH AND TO C.ELEGANS F42G9.8 |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - |
| CC | the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (see http://www.isb-sdb.ch/announce/ |
| CC | or send an email to license@isb-sdb.ch). |
| CC | ----- |
| DR | EMBL; Y09415; CAA70569.1; .. |
| DR | EMBL; AE000076; AAB91690.1; .. |
| KW | Modulation; Transferase; Plasmid. |
| SO | SEQUENCE 419 AA: 46569 MW: 8480C48E0416AAALF CXC64; |
| QY | Query Match 5.1%; Score 104.5; DB 1; Length 419; |
| Db | Best Local Similarity 21.5%; Pred. No. 0.15; |
| | Matches 78; Conservative 48; Mismatches 134; Indels 103; Gaps 18; |
| QY | 34 LSMKAQPERMAYLVSSWRSGSSFFVGOLFQHPDYFYLMEPAMHWMTPEKSTAMMLHMA 93 |
| Db | 1 MSRDVSPPLIFGLIGTRSGTTLAHLLOQHPPDTAPPE-----WL----- 43 |
| QY | 94 VFDLIRAVFLDMSVEDAYMEPPRQSSLPQWENSRALCSAPACDIIPODEII----- 147 |
| Db | 44 -----MLALFAFRVDRHHPA-----GASLIQAACBFLGRIDRISVRF 84 |
| QY | 148 -PRACRLLCSQPREVEYVEKACRSYHYLYKREFNLSQSLPLKDPSLNHYHLVND 206 |
| Db | 85 ADAASQYLAAGKRTLLDKTPRYV--WVLD-----YLHSLTP--EAP--HIL-LLEN 130 |
| QY | 207 PRAVRSERERTKGDIMDSRIYMGVEQKLKKEPDPPYMYQVLCOSQ-----LE 255 |
| Db | 131 PYAIAASLKSTMGVPFVESRCP-----PTVSQALAEIVTGTPTAAVALA 174 |
| QY | 256 IYKTIQSPLKALQER-----YLLVRDEDLARAFAVQTSR-----YFVGLLEPLPHLQTV 306 |
| Db | 175 LADVLGLPALAMQGRHHTQVYREERFERDEDIQVIAGLDGDPADVAGVEQT-- 232 |
| QY | 307 HNTTGKMGMD-HAFHTNARDALANYSQAMRWLSLPKXYSRLQKAGC-DAIMNLIGV----R 360 |
| Db | 233 -EYLRSSFGDRLLKKRAVDNRSV-ETWRTTLLTIEEMQTVYDVGADLIVELGIEQSIQ 290 |
| QY | 361 HVR 363 |
| Db | 291 HAR 293 |
| TS | RESULT 4 |
| ID | TPSB_CAEEL STANDARD. PRT: 359 AA. |
| AC | Q20351; |
| DT | 30-MAY-2000 (Rel. 39, Created) |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) |
| DE | Purative protein tyrosine sulfotransferase (EC 2.8.2.20) |
| DE | (Tyrosylprotein sulfotransferase) (TPST). |
| GN | F42G9.8. |
| OS | Caenorhabditis elegans. |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; |
| OC | Rhabditidae; Peloiderinae; Caenorhabditis. |
| OX | NCBI_TaxID=6239; |
| RA | SEQUENCE FROM N.A. |
| RA | STRAIN=Bristol N2; |
| RA | Taich A.; |
| RL | Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases. |
| CC | -1- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN |
| CC | ACIDIC MOTIFS OF POLYPEPTIDES (BY SIMILARITY). |
| CC | -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + protein tyrosine = |

| CC | adenosine 3',5',-bisphosphate + protein tyrosine-O-sulfate. |
|------|---|
| CC | -1- SUPCELLULAR LOCATION: Type II membrane protein. Colg1 membrane (By similarity). |
| CC | -1- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY. |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sdb.ch/announce/ or send an email to license@sdb-sdb.ch). |
| CC | EMBL, U00051; AAA91354.1; -. |
| DR | WormPep; F42908.8; CE07235. |
| DR | InterPro; IPR000863; Sulfotransferase. |
| DR | Pfam; PF00685; Sulfotransferase; 1. |
| CC | Hypothetical protein; Transferase; Transmembrane; Glycoprotein; Signal-anchor; 1. |
| KW | Signal-anchor; 1. |
| FT | DOMAIN 28 48 |
| FT | TRANSMEM 28 48 |
| FT | CYTOPLASMIC (POTENTIAL). |
| FT | SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL). |
| FT | LUMENAL, CATALYTIC (POTENTIAL). |
| FT | SEQUENCE 359 AA; 41469 MW; 9RC2F44539B8CC53 CRC64. |
| CC | Query Match |
| CC | Best Local Similarity 19.5%; Score 92; DB 1; Length 359; |
| CC | Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15; |
| QY | 52 RGSSEFVGGLREQQHDV---FYLMEPANHYMFMFKOSAWMLHAAVRLIAVFLCDMS 107 |
| DB | 95 RSGTILMRAILDADHDVDRGGFTMLLPSTLTQAQGRND-WNNNGI-----10E 143 |
| QY | 108 VFDAYMEGPFRQSSLSFQWENSRLACBAPCDIIPQ-DEIIPRACRLCSQOPEVEYER 166 |
| DB | 144 VFDDVAV-----SAFETITVAKHSELAR-----LCKNDP-----172 |
| QY | 167 ACRSSHYVLKVRFRFNQSLIPLLKDPSLNLIHYLVNDPRAVRRSRRTK-----GDLM 222 |
| DB | 173 ---TALMLPTIR-----RLTP-----NAKFTIMTIDAAVHSMITERRVPVAGVNT 216 |
| QY | 223 IDSRIVMGCHGKLRKEDQPYVMQVIGOSQ---LEIYKTIQSLPKALQERYLLVREYD 278 |
| DB | 217 SDEIMFQWQNELK-----MFFQCNNAAGQCIKYV-----YER 251 |
| QY | 279 LARAIVAOTSRMYEVVGLFELPHLOTWVHNITRGGMGGDHAN-TNARDALVNSQAWRMS 337 |
| DB | 252 LIQKAEIEILRTNLDLPLFSQQLRHODLIDDEVLDNDQFSASQVXNSITKALTSTWF 311 |
| QY | 338 LPY-EKVSRLQKAGDAMNLIGY 359 |
| DB | 312 DCFSEETLRKLDVAPFLGILGY 334 |
| CC | RESULT 5 |
| AC | ACF7_HUMAN STANDARD: PRT: 5430 AA. |
| AC | OSUPN3; OSUKP0; OSUC9; OSU540; OSU5033; |
| DT | 16-OCT-2001 (Rel. 40, Created) |
| DT | 16-OCT-2001 (Rel. 40, Last sequence update) |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) |
| DE | Actin cross-linking family protein 7 (Microphilin) (Tribeculin-alpha) (620 kDa actin-binding protein) (ABP620). |
| GN | ACF7 OR ABP620 OR KIAA0465 OR KIAA1251. |
| OS | Homo sapiens (Human). |
| CC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| CC | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. |
| NCBI | Taxid:9606; |
| RP | SEQUENCE FROM N.A. |
| FX | MEDLINE:20001959; PubMed:10529403. |
| RA | Okada T., Matsuda S., Nakatsuyawa S., Ichigotani Y., Iwashashi N., Takahashi M., Ishigaki T., Hamauchi M., |


```

Db 104 LGIAAPEPRRHVLLMATTRTSSGVEFNNQGNIFLYLEPLMHTERTVFERGANAVG 163
QY 92 MAV--KDLIAVAVLCMSVFDAMEGPRRO--SIFQWENSRAICSAACDITPO-DEI 146
Db 164 SALVYRDVLAQQLLCDLXYLESFISAPAEHHTALFRGGSSSLCEEPVC--TPSLKRV 221
QY 147 IPRAHCR-LLCSQOPEVEVEKACRSYSHVYKVEFFNFQSLYPLIKDPSLNLHLYALVR 205
Db 222 FEKYHCKNRGCGPLNTTLAEARRRKQHMALKTVRRLQLEFLDPLAEDRLDRIQTQVR 281
QY 206 DPAVRFRSRRRTGDLMDSRIVMGQHEOKLK-----KEDQPYVMQVYQCS-QLE 255
Db 282 DPAVAVVSR-----MV---AFSGKYESWKWMADEAPLQDE-VQRLRGCSIRLS 330
QY 256 IYKTIQSLPKALQERYLLVYREDLARAPVACSTSMYEFVGLLEPLPLQIWMVNHITRGKM 315
Db 331 AELGILN-PAWLGRLVLYEDVAPAPLKALEMRFAGIHPTPOVEKIRANTDAP-Q 388
QY 316 GDHAFHTNADALNVSQAMWSLPEYKVSRLQKACGDAMNLGVRHVRSEQGRNLLDL 375
Db 389 DSNQIYSTOKNSEOFEKWFPSIPFKLAQVYDACEPAMRLFGYKLASSAQELTNLSL 448
QY 376 L 376
Db 449 L 449

RESULT 2
YME7 MYCTU
ID YME7 MYCTU STANDARD: PRT: 388 AA.
AC 050695;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein RV2267c.
GN RV2267C OR MT2329 OR MTCY339.43.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC SRRAIN-H37R;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC SRRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gali J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z77163; CAB00968.1; -.
DR EMBL: AE007076; AA04611.1; -.
DR TIGR: MT2329; -.
DR Tuberculist; RV2267c; -.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Hypothetical protein; Complete proteome.
SQ
Sequence 388 AA; 46062 MW; 5DSD0263275a9B24 CRC64;

Query Match
Best Local Similarity 19.3%; Pred. No. 0.014;
Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

QY 45 VYLSMSRSGSSFFVQOLF---GQH--PDVEYLMFPAHVMYMTFKGSTAMLMHNAVRLIR 99
Db 83 IFVGHWRGTGTLHLLHLLVYDDRHGTGTECLAB--HHFLD-----LW----- 125
QY 100 AVFLCDMSVFDAMVEGPRRQSSLFQWENSRAICSAACDITPO-DEI-----PRAHC 152
Db 126 -----FAHYE-----FLYSKHRAMQMDLSLHHQDEDFWCMQGLPSYL 167
QY 153 RLCSQOPEVEVEKACRSYSHVYKVEFFNFQSLY-----PLKDPSSLNLI 200
Db 168 TAFPRRPQYEEYLDLDQVAPRELE--IKRITLRFVQVYFRRKTVILKNPHTSRI 225
QY 201 -----VHLVDPRAVFRSRRRTGDLMDSRIVMGQHEOKLKEDQPY--VM 246
Db 226 KYLLEVFQAKTIHLYRDPYVY-----PSITHLKALYRHGLQQPFDDLD 273
QY 247 QVYCSQLEIYVTIQLSPKALQ-ERYLLVYREDLARAPVACSTSMYEFVGLF---ELPH 301
Db 274 DVVSYVYVLYRKLDLDEGRFLVDPTRFEYELREDLIGDEGLRLRYQLIGLDEFCYLP 333
QY 302 LQTVWNIHNRGKMGMDHA-FHTNARDALNVSQAMWSLPEYKVSRLQKACGDAMNLGVR 359
Db 334 LKQY-----LADHDYKTN-----YQLVDEKRALVDEHWGEITIDRIGTD 373
QY 360 RH 361
Db 374 RH 375

RESULT 3
NOEE RHISN
ID NOEE RHISN STANDARD: PRT: 419 AA.
AC P55472; P72326;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Nucleation protein noee (EC 2.8.2.-).
GN NOEE OR Y4HB.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym PNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97361801; PubMed=9218762;
RA Hann M., Jabouri S., Quesada-Vlencens S., Freiberg C., Perret X.,
RA Broughton W.J., Fellay R.;
RT "Supination of Rhizobium sp. NGR234 Nod factors is dependent on noee,
RT a new host-specificity gene."
RL Mol. Microbiol. 24:1119-1129(1997).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes."
RL Nature 387:394-401(1997).

```


RESULT 15

C69679

polyketide synthase pksM - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Nov-2000

C:Accession: C69679

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Berber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerston, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holstappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Mathers, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
Y., M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A.; Authors: Schleicher, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot
akuch, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MIMD:98044053; PMID:9384577

A:Accession: C69679

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-4273 <KUN>

A:Cross-references: GB:Z99113; GB:AL009126; NID:q2634090; PIDN:CA13603.1; PID:q2634103

A:Experimental source: strain 168

C:Genetics:

A:Gene: pksM

C:Superfamily: Bacillus subtilis polyketide synthase pksM; 3-oxoacyl-[acyl]-carrier-prote

C:Keywords: carrier protein

F:293-363/Domain: acyl carrier protein homology <ACPI>

F:414-829/Domain: 3-oxoacyl-[acyl]-carrier-protein synthase I homology <OAS1>

F:2340-2732/Domain: 3-oxoacyl-[acyl]-carrier-protein synthase I homology <OAS>

F:3145-3320/Domain: short-chain alcohol dehydrogenase homology <SAD1>

F:3550-3942/Domain: 3-oxoacyl-[acyl]-carrier-protein synthase I homology <OAS3>

F:4138-4208/Domain: acyl carrier protein homology <ACPI>

Query Match

Best Local Similarity 4.4%; Score 89.5; DB 2; Length 4273;

Matches 46; Conservative 34; Mismatches 87; Indels 41; Gaps 10;

QY 161 FEVEKACRSYSHVLEKREFFNQLSLPLKDPSP---LNLIIVHLVRDPAVFRSRR 216

DB 36 YQVQEHGVAQGELEK-----NLTIFFPLIADSEYDIALTIHVSERESEGTWSIIIDGOK 90

QY 217 TKGDLMTDSRIWQGH-QKLKEDDPYIYQVYCQSELEITYKTIOSLPKALDERILVR 275

DB 91 QHGESLSDKR--QYETADMHRKEGTAFASFI---DLNQKSTADRIINLDEIYEOCR 142

QY 276 YEDLARAPVQOT-SRME-----FVGLLEFLPHQTWVHNIT--RKGMDHAFTN 323

DB 143 SQELVHTGMKAKESQITEAKGAVIDLAVQGFALRHSDAFLPHPTLIDSGIGS----- 196

QY 324 ARDALNVQAARMSLPYKVS---RIQKAC 350

DB 197 --SCLISDQTMVLPFLYSEFSASERLQKGC 224

Search completed: January 11, 2003, 01:21:55
Job time : 54 secs

This Page Blank (uspto)

Qy 340 YKYSRIQKACGDAMNLGYHVRSEQ-----EQRNLLDLSTWTVPEQI 385
 Db 1344 KLVYSRASYNGEYARALSTYILEGDEKSORLLEQFTFIVEGSLRDPDSV 1398

RESULT 9

T16350

hypothetical protein F42G9.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000

C:Accession: T16350

R:Taich, A.

submitted to the EMBL Data Library, March 1996

A:Description: The sequence of C. elegans cosmid F42G9.

A:Reference number: Z18498

A:Accession: T16350

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-359 <TAI>

A:Cross-references: EMBL:000051; NID:g1216305; PID:g1216308; PIDN:AAA91354.1; CESP:F42G9

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F42G9.8

A:Mutons: 26/2; 50/2; 113/2; 155/3; 220/3; 290/2

C:Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8

Query Match

Best Local Similarity 4.5%; Score 92; DB 2; Length 359;

Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;

Qy 52 RSGSSTVGFQGHDPV---FLIMEPAHVMKTFKOSTAMLMHMAVRDILRAVFLCDMS 107

Db 95 RSGTTLRALIDAHDPVRCGCGETMLPSFLTQAGWRND-WVNNSGI-----TQE 143

Qy 108 VFDAVMEPPGRQSSLEFQWENSRLCSAPACDIIPQ-DEIIPRAHCLLCSQPFVEVER 166

Db 144 VFDDAV-----SAFTTEIYVAHSEIAPR-----LCNKDP----- 172

Qy 167 ACSRSYHVLKEVFEFNLQSLYELKDPSLNLIHVLVPRPAVFRSRETK-----GDLM 222

Db 173 ---YALMLPTIR-----RLYP-----NAKFIIMIRDAVHSMIERKVPVAGYNT 216

Qy 223 IDSRIVGQHEOKLKKEDQYVYMOYICQO-----LEIKTISLKLAKOERLLVRYED 278

Db 217 SDEISMVQNMQLERK-----MTPOCNNAPEQCikv-----YER 251

Qy 279 LARAFAQTSRMVEFVGLFELPILQTVHNTTRGKMGCDHAFH-TNARDALNVQAMRMS 337

Db 252 LIQKPAEEIIRITNFDLPFSQOMLRHODLIGDEVLDNQEPSASQVKNISMTKALTSWF 311

Qy 338 LPY-EKYSRIQKACGDAMNLGY 359

Db 312 DCFSEETLRKLDVAPFLIGLY 334

RESULT 10

JEO196

hydroxysteroid sulfotransferase (EC 2.8.-.-) - mouse

C:Species: Mus musculus (house mouse)

C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 18-Jun-1999

C:Accession: JEO196

R:Sakakibara, Y.; Yanagisawa, K.; Takami, Y.; Nakayama, T.; Saito, M.; Liu, M.C.

Biochem. Biophys. Res. Commun. 247, 681-686, 1998

A>Title: Molecular cloning, expression, and functional characterization of novel mouse s

A:Reference number: JEO196; M01D:98521187; PMID:9647753

A:Accession: JEO196

A:Molecule type: mRNA

A:Residues: 1-338 <SAK>

A:Cross-references: GB:AF026072; NID:q2570897; PIDN:AA069918.1; PID:q2570898

C:Comment: This enzyme catalyzes the transfer of a sulfonate group from the active sulfat

C:Superfamily: alcohol sulfotransferase

C:Keywords: transferase

Query Match 4.5%; Score 91.5; DB 2; Length 338;
 Best Local Similarity 22.2%; Pred. No. 3.9;
 Matches 52; Conservative 32; Mismatches 79; Indels 71; Gaps 11;

Qy 67 DVEYLMEPAHVMKTFKOSTAMLMHMAVRDILRAVFLCDMSVFDVAMEPPGRQSSLCQM 126

Db 59 DIFIVYIP-----KSGTMMWIE-----IVCLIKND-----GDSWIRSEPIWO- 96

Qy 127 ENSRALCSAPACD-IIPODEIIPRAHCLLCSQPFVEVERKACRSYHVLKEVFPNLQ 185

Db 97 -----RAPWCETIISAFNVLDPSPRIMSHPLIELETKA-----FF--- 133

Qy 186 SLVPLKDSNLNLIHVLVPRPAVFRSRETKQDMLDSRIVMGQHQKIKKEDQPIYV 245

Db 134 -----SSKAKVLYVGNRPDVVVS-----LYYYSKILA-QQ-----LKDPTMPQOF 172

Qy 246 MQVICQSLIEIKTIQSLKAL-----QERYLLVREYEDLARAPVQTSRMVEFV 295

Db 173 LQNFLKGEVQGSWFDHITKQWIRMONQENFLPITVEELQDQDLRGSVQNICERLG 226

RESULT 11

S11737

resistance protein Mx3, interferon-regulated - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Mar-2000

C:Accession: S11737

R:Meier, E.; Kunz, G.; Haller, O.; Arnheiter, H.

submitted to the EMBL Data Library, April 1990

A:Description: Unexpected activity of rat Mx proteins against a Rhadovirus.

A:Reference number: S11735

A:Accession: S11737

A:Molecule type: mRNA

A:Residues: 1-659 <EMBL>

A:Cross-references: EMBL:X52713; NID:q56724; PIDN:CAA36937.1; PID:q56725

C:Superfamily: dynamin-related protein vps1

Query Match

Best Local Similarity 4.5%; Score 91; DB 2; Length 659;

Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;

Qy 62 FQGHDPVYLMPEPAHVMKTFKOSTAMLMHMAVRDIL---RAVFLCDMSVFDVAME-- 114

Db 301 FKHPQFALLLEDG-----KATVPCLAERTLMELISHICKSLPLENQIKESQSTSE 353

Qy 115 -----PPGRQSSLFQWENSRLCSAPACDIIPQDEIIPRAHCLLCSQPF--- 161

Db 354 ELQKYGADIPEDENKTLFLEIKINAF-NODITAVEGEELVREKRECHLFTLKKKEFFLV 412

Qy 162 -EVEKACRSYSHVLKEVFEFNLQ-----SLVPLKDSNLNLI 200

Db 413 SEEIRNFQKSDALYKEVTFEMQYRGRELPGFVNYKTFENIIRQIKTLEBAMEN-- 470

Qy 201 VHLVND-PRAYR-----SRETKGDLMDISIVMGQHQKIKKEDQPIYVQV 248

Db 471 LHKVEIYRAAFVTSSEKNESEFNLHRTTKSKLEDLRL--EDETAEKALRLHFMQ 527

Qy 249 ICOSQLEIYKTIQSLKALOEKRYLLVREYEDL-----ARAPVQTSRMVEFVGL 296

Db 528 ILYQDDQIYR-----KALQK-----VREEAEDEERKHKGRKRSQSNLQTSMD----- 572

Qy 297 EFLPHLQTV---VHN 308

Db 573 EIFQHLNAYROEAAH 587

RESULT 12

E64477

replication factor C homolog - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: E64477

14 VS0MAL1A1EPMHYSHN1SS1SMKQAPENMHV1LSSWSSGSPFGOLFGCHPDVFLYME 73

Db 990 VFKAGIFENFETGSHSLKANTLLTKIR-----ETGIEIPQGLFENHPILALAE 1040

QY 74 PAMVWMTFKOSTAMLMMAVVDLIRAVFLCDMSVFDAYMEPGPRQSSLFQWENSRAIC 133

Db 1041 EADH-----RESKAF-----AVIEPAKQGHYPLHMSSSEHS 1073

QY 134 SAPA-----CDIIPDEIT-----PRANCILCS 157

Db 1074 SASSMRESAIIHASSNSGEFRYSKAGARISGINPTFVETISFVLENSPRKIH-VCV 1132

QY 158 QQPFVYVKACRSYSHVYLKEVRFENLOSILYPLKDPSEL-----NLHY-----HLVRDPR 208

Db 1133 DFNIMIRGCRSDDAIMASFRTFDLAKA-PLFRIGLLGLEENRMMLFDMHNLISGV 1191

QY 209 AVFRSRETKDGLMID--SRIVMGHEGKLKKEQDPYVMQVIOQSOLEIKTTIOSLPKA 266

Db 1192 SI-----GIMLEULARIKYGQPLDLRLQKDYAVMG--SRQAGEYKKQDAYWKE 1240

QY 267 L--QERYLLVREDIARAPV 284

Db 1241 VFAGELPVLQILSDYPRPV 1260

RESULT 8

113288

mei-41 protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: 113288

R:Hard, K.L.; Santerre, A.; Sekelsky, J.J.; McKim, K.S.; Boyd, J.B.; Hawley, R.S. Cell 82, 815-821, 1995

A:Title: The mei-41 gene of D. melanogaster is a structural and functional homolog of A:Reference number: 211072; MUID:95401271; PMID:7671309

A:Accession: 113288

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2354 <HAR>

A:Cross-references: EMBL:U034925; NID:g998351; PID:g998353; PIDN:AA046881.1

C:Genetics:

A:Gene: mei-41

A:Cross-references: FlyBase:FBgn0004367

A:Introns: 650/3; 748/3; 2313/3

C:Function:

A:Description: involved in cell cycle checkpoint and meiotic recombination

Query Match 4.5%; Score 92.5; DB 2; Length 2354;

Best Local Similarity 20.5%; Pred. No. 43;

Matches 85; Conservative 52; Mismatches 147; Indels 131; Gaps 20;

QY 28 SHNHS---LSMKAQGERMHVLYLSWRSGSSFVGQLFGQHPDVFYLMERAHVWMTFRQ 84

Db 1058 NYFNASPDHPLNLISSDDPAVALALISLCRGYQF--QQNTKHYDSFSL----- 1102

QY 85 STAMLMMAVVDLIRAVFLCDMSVFDAYMEPGPRQSSLFQWENSRAICSAAPCIIPOD 144

Db 1103 -----ALGETIA-----LCGIS-----PREKKYQWLQW-----SLPA-----RMQ 1133

QY 145 EIIIRAHCHLLCSQGFEEVYERAC--RSYSHVYLKEVRF-----NLOS 186

Db 1134 LMEPMIHSCTVCVHRPSTCQQOPLFSGHSHNYEEMAFIMASRLIDYIPSSGKRHLISS 1193

QY 187 LVPFLKLD-----PSLNLHY-----HIVRDAVERFRSRETKDGLMIDSRIV 228

Db 1194 YKPCIKRDSNMILSTFYPYILLHLLCTTEQRNHIIDEEFMAVLOANES-----SSSV 1246

QY 229 MGQHEKLLKKEQPYVMQVIOQSOLEIKTTIOSLPKALQERYLLVREDIARAPVQTS 288

Db 1247 RGHQELGAIKEN-----AFKQESKRYANGI-KPLASTLVSRKEDSSRP----- 1291

QY 289 RMVEFGLDFLEPIQLQTW-----VANIIRGKGNGDAHFHTNAR--DALVNSQAWMSLP 339

Db 1292 RLAGKLCALDLPLQMLRWRWRIHGRSTG-CGPPETIDSNYKKIHFEFINL-----IP 1343

Db 389 DSNRTISQKNSSEQFEKMFSTFLKAYQVQDACEPAKRLFGTKLASSAQEULTNLSL 448
 QY 376 L 376
 Db 449 L 449

RESULT 5

E95934
 C:Species: *Sinorhizobium meliloti*
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E95934
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21395508; PMID:11481431
 A:Accession: E95934
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CAC49141.1; PID:g15140626; GSPDB:GN00167
 A:Experimental source: Strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pel, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebut, P.; Vandenol, M.; Vorholter, F.J.; Weidner, S.; Welis, D.H.; Wong, K.; Yen, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMD21237
 A:Genome: plasmid

Query Match 8.2%; Score 168; DB 2; Length 307;
 Best Local Similarity 22.7%; Pred. No. 9.2e-07;
 Matches 83; Conservative 50; Mismatches 131; Indels 102; Gaps 16;

QY 36 MKAPERNHVLVLSWRSGSSFFVQGLFGQHPDYF---YLMERPAMHVMTEFKOSTAMLMH 92
 Db 1 MPSGPVATAYIAGYGRSSGTTLDIALGCHAAVYAGCEITSLTRHWRENEYCA----- 53
 QY 93 AVRDILRAVELCOMSVFADAYMEBGRPROSSLFQWENSRLCSAPACDIIPODETIIPRANC 152
 Db 54 -----CGNATRCDSFMSVRR-----FMSDG-----QDPGLMEYXC 84
 QY 153 RLCSQOPEFVEVKACR-----SYSHVYLKEVRFN-IQS----- 186
 Db 85 AL---QQKFEGLSMATLTLSGMLGKQFSLYILHTRKLFESAMOSCGRGVYVDSKLPGR 141
 QY 187 LYPLLKPSLNLHIVLVRDPRAVFRS-----RERTKGDLMTDSRIYMGHEQKLKED 240
 Db 142 AMVAQAIPGIDMVIHIVLRDGRGAVMSLLKGYERDAKSG-----IQKEL 185
 QY 241 QPIYVWVYVQSOLEIKTKTQSLPKAL-QERYLLRYEDLARPVPVQTSRMEFFVLEFL 299
 Db 186 KPSVSEFTALRWSV-VLAVEVLSRKISGSKVARKVREDFASDPVAVMOQIGTFLELD-L 243
 QY 300 PHLOTWVHNITRGKMG-----DHAFTNARDALNVSOAMRSLYEKVSRLQKACGDA 353
 Db 244 SQTSTSEN---GEAMPGHQVAGNRLKMASTALKKDETRTRMPARQCVSFQRLGGM 300
 QY 354 MNLGY 359
 Db 301 LRRGY 306
 RESULT 6
 G70729
 hypothetical protein RV2267c - *Mycobacterium tuberculosis* (strain H37Rv)

C:Species: *Mycobacterium tuberculosis*
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: G70729
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
 A:Reference number: A70500; MUID:96295987; PMID:9634230
 A:Accession: G70729
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-388 <COL>
 A:Cross-references: GB:Z77163; GB:AL123456; NID:93261610; PIDN:CAB00968.1; PID:e25507
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: RV2267c

Query Match 5.7%; Score 116; DB 2; Length 388;
 Best Local Similarity 19.3%; Pred. No. 0.037;
 Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

QY 45 VLVLSWRSGSSFFVQGLF---GQH---PDVYLMERPAMHVMTEFKOSTAMLMAYRDLIR 99
 Db 83 IFVGHMRTGTLHLHLVVDNRHTGPTGECLEP--HFFLD---TEW----- 125
 QY 100 AVFLCMSVFADAYMEBGRPROSSLFQWENSRLCSAPACDIIPO-DELI-----PRAHC 152
 Db 126 -----FABYVE-----FLVSKHRMMDNLSLHHPODERFVWCWGGLPSYL 167
 QY 153 RLCSQOPEFVEVKACR---SYSHVYLKEVRFN-IQS-----PLKDPSTLNI 200
 Db 168 TIAPRPRPQVEEYLDLEQVARELE--IKRRLFRFQVQYVFRKRVILKNPHSRI 225
 QY 201 -----VHLVRPRAVFRSRERKTDMLDSRIYMGHEQKLKEDQPY--VM 246
 Db 226 KVLVEFPOAKFIHVRDPRVYVY-----PSTIHLKALYRHLGQDPFDLD 273
 QY 247 QVICSOLEIYTKTQSLPKALO-ERYLLRYEDLARAFAQTSRMEFFVLE---FLPH 301
 Db 274 KQVSTVYVDLYKRLDGRGLVPTFTELYRVEDLLGDEGQLRKLXQHLGLGDEFECYLR 333
 QY 302 LOTWVHNITRGKMGDHA-FHTNARDALNVSOAMRSLYEKVSRLQKACGDANMLGY- 359
 Db 334 LQY-----LADHADYKTN-----YQLTVEQRAIVDEHWGHIIDRYGD 373
 QY 360 RH 361
 Db 374 RH 375

RESULT 7

I40485
 surfactin synthetase component I - *Bacillus subtilis*
 N:Alternate names: competence protein srfA; surfactin production protein srfA; surf
 C:Species: *Bacillus subtilis*
 C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000
 C:Accession: I40485; B69718; S3517; A37323; S45967; A43705; S34985
 R:Cosmin, P.; Rodriguez, F.; de Ferra, F.; Graedel, G.; Perego, M.; Venema, G.; Van S
 M.O. Microbiol. 8, 821-831, 1993
 A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthe
 A:Reference number: I40485; MUID:93360813; PMID:8355603
 A:Accession: I40485
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-358 <RES>
 A:Cross-references: EMBL:X70356; NID:9396480; PIDN:CAA49816.1; PID:g396481
 A:Experimental source: strain M168 derivative of JH642
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brulliet, S.; Brusch, C.V.; Calzavara, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997

JC7351
 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
 C:Species: Homo sapiens (man)
 C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
 C:Accession: J07351
 R:Uchiyama, K.; Fasaki, F.; Kadonaga, K.; Matsukawa, T.; Kurosawa, N.;
 Biochem. Biophys. Res. Commun. 274, 291-296, 2000
 A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
 A:Reference number: J07350; M01D:20374462; PMID:10913333
 A:Accession: J07351
 A:Molecule type: mRNA
 A:Residues: 1-486 <UCH>
 A:Cross-references: DBJ:AB040711
 C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylated
 sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
 C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.8%; Score 606.5; DB 2; Length 486;
 Best Local Similarity 37.0%; Pred. No. 4e-44;
 Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERNHVLYSSMRSGSSFFVQLFGQHPDVFYLMERAHVMTFKOSTAMLMHAYDLIRA 100
 DB 100 EKQHIYVHATWRTGSSFFGLFENQHPDVFYLEPMWHLMQALYPGDASTLQGLRDLMS 159
 QY 101 VELCDMSVFDAYMEPS-PRQ-----SFLQWENSRALCSAPACDIIPODE---I 146
 DB 160 LRFCDSESVLRVAPGDPAPAPADTANLTALFRWRTKAVCSPLCGAPARAEVGL 219
 QY 147 IPRAGRLCSQGFEEVERKACRSYSHVYLKEVFPNLSYPLKDPSLNHIYLVDRD 206
 DB 220 VEDTAEKRCPPVPAIRALAEKCRKYPVYVVKIDVRLDLCVLPRLDPGLNKKVOLFPRD 279
 QY 207 PRAVFRSRETKGDLMDIS-----RIYV-----GQHQKLRKEQPY 244
 DB 280 PRAVHNSRLKSGGLRSTQVLTQRQDRFRHRYLLAAGVARGPQSGRALPAPRADE 339
 QY 245 ---VMQVICSQGLTYKTQSLPKALOERYLLVREYEDLARPAVQTSRMVEFGIEFLP 300
 DB 340 FLTGAEIVICEALRDLLEFARGAPAMLRRLRYRLREYEDLPVRAOIRLRLFFSGRALA 399
 QY 301 HLTQWHTNRGKMGV-DHAFHTNARDALNVSGAWMSLPEYKVSRLQKACGDAMMLGY 359
 DB 400 ALDAFALNMTGKAGYADRPFLHSADARAVAHNRERLSREVRGEVACAPAMLLAY 459
 QY 360 RHVRSDEQ 367
 DB 460 --PRSGEE 465

RESULT 3
 J070261
 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
 C:Species: Homo sapiens (man)
 C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
 C:Accession: J070261
 R:Uchiyama, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka,
 T.
 J. Biochem. 124, 670-678, 1998
 A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
 A:Reference number: J070261; M01D:98391845; PMID:9722682
 A:Accession: J070261
 A:Molecule type: mRNA
 A:Residues: 1-484 <UCH>
 A:Cross-references: DBJ:AB014679
 C:Comment: This protein catalyzes the transfer of sulfur from 3'-phosphoadenosine
 C:Superfamily: chondroitin 6-sulfotransferase
 C:Keywords: sulfotransferase

Query Match 28.9%; Score 588.5; DB 2; Length 484;
 Best Local Similarity 36.0%; Pred. No. 1.4e-42;
 Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERNHVLYSSMRSGSSFFVQLFGQHPDVFYLMERAHVMTFKOSTAMLMHAYDLIRA 100
 DB 117 KRWNVYVTTIRSSSSFFGLFENQHPDVFYLEPMWHLMQALYPGDASTLQGLRDLMS 176
 QY 101 VELCDMSVFDAYMEPS-PRQ-----SFLQWENSRALCSAPACDIIPODEIIPRACRL 155
 DB 177 LYRCDLSFQILSPAGSGRNLTLTGFGATNKVCCSPICAPV-KREVGLVDKVC 235
 QY 156 -CSQGFEEVERKACRSYSHVYLKEVFPNLSYPLKDPSLNHIYLVDRPAPFRSR 214
 DB 236 KCPQRLARFEEDCRKRYTLVKGVRFDVAVALPRLDPAIDLKIVHVPDPAVASSR 295
 QY 215 ERTKSDMLDRIRYV-----GQHQKLRKE---DQPY---VMQVICSQ 252
 DB 296 IRSRGLRSTQVLRSDPPAHKRPFLDAAAGHKLGAKEBSVGSPADYHALGAEVLCNS 355
 QY 253 QLEIKTQSLPKALOERYLLVREYEDLARPAVQTSRMVEFGIEFLPHLTQWHTNRG 312
 DB 356 MAKTQIALQPPDMLQGHYLVREYEDLVGEVPTLRRYVDEGLVSPEDMQLNMTSG 415
 QY 313 KQMGDAHTNARDALNVSGAWMSLPEYKVSRLQKACGDAMMLGVRHRSDEQRLN 372
 DB 416 SGSSSKPPVVSARNAQTQANAMRTALFQQLQVVEECYQPMAYIGERVNSPEEVKDL 475
 QY 373 IDLL 376
 DB 476 KTL 479

RESULT 4
 A57397
 chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000
 C:Accession: A57397
 R:Fukuta, M.; Uchiyama, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuc,
 J. Biol. Chem. 270, 18575-18580, 1995
 A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransferase
 A:Reference number: A57397; M01D:95355490; PMID:7629189
 A:Accession: A57397
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-458 <FUK>
 A:Cross-references: GB:D4915; NID:9711262; PIDN:BA08655.1; PID:9711263
 C:Superfamily: chondroitin 6-sulfotransferase
 C:Keywords: sulfotransferase

Query Match 24.6%; Score 500.5; DB 2; Length 458;
 Best Local Similarity 33.2%; Pred. No. 4.6e-35;
 Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

QY 34 LSMKAQPERMHVLYSSMRSGSSFFVQLFGQHPDVFYLMERAHV-MTFKOSTAMLMH 91
 DB 104 LGIAPEPRRHVLLMATRTGSSFFGEFFNOGNIFFLEFLMHTERTVTEPGGANAVG 163
 QY 92 MAV--RDLIRAVFLCDMSVFDAYMEPS-PRQ--SFLQWENSRALCSAPACDIIPO-DEI 146
 DB 164 SALVVRDVLQQLDLDLTIESFFISPAEHEHTALFRGSSSHSCSEPV--TPSEKVV 221
 QY 147 IPRACR-LLCSQGFEEVERKACRSYSHVYLKEVFPNLSYPLKDPSLNHIYLVDR 205
 DB 222 FEKYCKNRKRGPLNLTIAEACRKKQMALKYAIRLEFLQPLADPRDLRIQVLR 281
 QY 206 DPRAVFRSRETKGDLMDISRLVMGQHQKLR-----KEDQPYVMQVICSQ-QL 255
 DB 282 DPRAVLVR-----NV--AFSGKYSMKWAAEGCAPLQDEE-VORLNGNESTRLS 330
 QY 256 IYKTQSLPKALOERYLLVREYEDLARPAVQTSRMVEFGIEFLPHLTQWHTNRGKM 315
 DB 331 AFLGLRQ-PRWLRGKRYMLVREYEDVARAPLRKALEMYRRAGHLPITQVDEWLRANTQAR-Q 388
 QY 316 GDHAFHTNARDALNVSGAWMSLPEYKVSRLQKACGDAMMLGVRHRSDEQRLNLDL 375

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 11, 2003, 01:18:46 ; Search time 40 Seconds

(without alignments)
927.697 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038
Sequence: 1 MLPPKMKLLLEVSQMAIL.....EQRLNLLDLSWTWPEQIR 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224.

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-----------------------------|
| 1 | 623 | 30.6 | 484 | 2 | JC7350 N-acetylglucosamin |
| 2 | 606.5 | 29.8 | 486 | 2 | JC7351 N-acetylglucosamin |
| 3 | 588.5 | 28.9 | 484 | 2 | JC7351 N-acetylglucosamin |
| 4 | 500.5 | 24.6 | 458 | 2 | JE0261 Chondroitin 6-sulf |
| 5 | 168 | 8.2 | 307 | 2 | E53934 Probable enzyme, C |
| 6 | 116 | 5.7 | 388 | 2 | G70729 Probable enzyme, C |
| 7 | 95 | 4.7 | 358 | 2 | G70729 Probable enzyme, C |
| 8 | 92.5 | 4.5 | 2354 | 2 | T13288 surfactin syntheta |
| 9 | 92 | 4.5 | 359 | 2 | T16350 met-41 protein - f |
| 10 | 91.5 | 4.5 | 338 | 2 | JE0196 hypothetical prote |
| 11 | 91 | 4.5 | 659 | 2 | S11737 hydroxysteroid sul |
| 12 | 91 | 4.5 | 644 | 2 | E64477 resistance protein |
| 13 | 89.5 | 4.4 | 346 | 2 | E86319 replicational factor |
| 14 | 89.5 | 4.4 | 504 | 2 | T07120 probable flavonol |
| 15 | 89.5 | 4.4 | 4273 | 2 | C69679 polyketide synthas |
| 16 | 89 | 4.4 | 524 | 1 | A36205 mitochondrial proc |
| 17 | 89 | 4.4 | 1140 | 2 | G71335 probable transcrip |
| 18 | 89 | 4.4 | 1568 | 1 | P00093 genome polyprotein |
| 19 | 88 | 4.3 | 359 | 2 | T47187 hypothetical prote |
| 20 | 88 | 4.3 | 466 | 2 | E84132 aminopeptidase B13 |
| 21 | 87.5 | 4.3 | 380 | 2 | T42755 tyrosylprotein sul |
| 22 | 87.5 | 4.3 | 4063 | 2 | T42993 probable spectrin |
| 23 | 87.5 | 4.3 | 4101 | 2 | T23630 hypothetical prote |
| 24 | 87 | 4.3 | 2500 | 2 | G88493 protein F57B9.2 [1 |
| 25 | 86.5 | 4.2 | 748 | 1 | S08680 methylmalonyl-CoA |
| 26 | 86.5 | 4.2 | 831 | 2 | S26675 DNA-directed DNA p |
| 27 | 86.5 | 4.2 | 2352 | 2 | T06077 splicing factor PR |
| 28 | 86 | 4.2 | 320 | 2 | A40216 flavonol 4'-sulfot |
| 29 | 86 | 4.2 | 823 | 2 | S48986 probable protein k |

| | | | | | |
|----|------|-----|------|---|---------------------------|
| 30 | 86 | 4.2 | 869 | 2 | A56558 probable protein k |
| 31 | 85.5 | 4.2 | 344 | 2 | H83928 hypothetical prote |
| 32 | 85.5 | 4.2 | 764 | 1 | S14113 1-phosphatidylinos |
| 33 | 85 | 4.2 | 632 | 2 | T18692 hypothetical prote |
| 34 | 85 | 4.2 | 672 | 2 | T34896 probable DNA recom |
| 35 | 85 | 4.2 | 1237 | 2 | T17529 hypothetical prote |
| 36 | 85 | 4.2 | 2688 | 2 | T49477 alpha-A-crystallin |
| 37 | 84.5 | 4.1 | 217 | 2 | J01358 Cl protein - Misco |
| 38 | 84.5 | 4.1 | 646 | 2 | G86853 threonine-tRNA lig |
| 39 | 84.5 | 4.1 | 1031 | 2 | T43458 hypothetical prote |
| 40 | 84.5 | 4.1 | 2493 | 2 | S45734 probable membrane |
| 41 | 84 | 4.1 | 390 | 2 | AB2337 group 2 sigma 70-t |
| 42 | 84 | 4.1 | 1189 | 2 | T42726 guanine nucleotide |
| 43 | 84 | 4.1 | 1217 | 2 | T00270 hypothetical prote |
| 44 | 84 | 4.1 | 1583 | 2 | T14176 probable phosphati |
| 45 | 83.5 | 4.1 | 518 | 2 | E84536 hypothetical prote |

ALIGNMENTS

RESULT 1

JC7350 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse

C:Species: Mus musculus (house mouse)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000

C:Accession: JC7350

R:Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, Biochem. Biophys. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of A:Reference number: JC7350; MUID:20374462; PMID:10913333

A:Accession: JC7350

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB040710

C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosyl sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand

C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 30.6%, Score 623, DB 2, Length 484;

Best Local Similarity 34.2%, Pred. No. 1,56-45;

Matches 136; Conservative 65; Mismatches 151; Indels 46; Gaps 7;

| | | | |
|----|-----|--|-----|
| QY | 30 | NISSLMKAOPEPMHVLSSWSSGSSFFVQIFGHPDVFILMEPAHVMFTKOSTAM | 89 |
| DB | 87 | NLSAAGEAVTQEKHIIYHATMRTGSSFLGELFNHPDVFILMEPMHLMQALPGDAES | 146 |
| QY | 90 | LMHAYRDLIRAVFLCDMSVFDAYMEGPRRQ-----SLPQWENSALCSAPACD | 139 |
| DB | 147 | LOGALRDMILRSFLRCDFSVLRLYAQPGDGERAPDSANLTYAMLFRTNIVICSPPICP | 206 |
| QY | 140 | IIPDQ-----IIPRAHCHULCSQGFVEYKACRSYSHVLEKEYRFFELQSLYPLKDP | 195 |
| DB | 207 | AAPRRADVGLVEDKACSTCEPVSRLALEACRKIPYVIVIKDVLDELGLVLELRDPE | 266 |
| QY | 196 | LNLIHVLHVDPRFAVRSRERTKGLMDS-----RIYV-----GOHE | 233 |
| DB | 267 | LNKLVQFLRDBRAVHNSRLKRSQGLRESIQVLTFRQGDHFHVLAAHGVADARPQGA | 326 |
| QY | 234 | OKLKEDEPYV---VMQVIOQSLEIKTIQSLPKALOEFYLLVREEDLARAIVAQTSR | 289 |
| DB | 327 | RALPAPRAPDFLTSALVEICAWLRDLFTRGAPAMLRRTLRKREEDLWQDQDARR | 386 |
| QY | 290 | MEEFGLFEFLPHLOTWVINITRGKMG-DHAFTNARDALNVAQWMSLYEKVSRLQK | 348 |
| DB | 387 | LLRFGLFTFLAALDAFANMTRGSAYGADRPFLHSARARAVHVMRRLSQEDVQVET | 446 |
| QY | 349 | ACGDAMLLGTRHNSDEQORNL-----LDLSWTWY | 361 |
| DB | 447 | ACAPAMRLAYPRSGDERDKRTVREGTEPLETKAMAV | 484 |

RESULT 2

CC during transplantation.
CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.
XX
SQ Sequence 390 AA:
Query Match 49.5%; Score 1008; DB 22; Length 390;
Best Local Similarity 52.3%; Pred. No. 3,7e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
QY 1 MLPRKMKLLFLVSOAILALFFHMYSHNSSL.SMKAPERMHVLYLSSWSSGSSFFVQ 60
Db 14 LLLAQTCLLFLITSRP-----GSSPAGGEDRHHVLYLSSWSSGSSFFLQ 59
QY 61 LFGQHPDVFYLMPEPAHVMWTFKOSTAMMLHNAVRLIRAVFLCDMSVFDAYMEPPRRQ 120
Db 60 LFSQHPDVFYLMPEPAHVMWTFKOSTAMMLHNAVRLIRAVFLCDMSVFDAYM-PQSRNL 118
QY 121 SSLFOWENSRALCSAPACDIIPODEIIPRAHCRLLCSQOFVEVEKACRSYSHVLYKEVR 180
Db 119 SAFENMATSRLCSPACSAFPRGTISKQDYCKTCTROPFSILARECYSYSHVLYKEVR 178
QY 181 FENLQSLYPLKDPSSLNLHIVLRDPAVERFSRERTKCDLMDISRIYWGQHEQKLRKD 240
Db 179 FENLQSLYPLKDPSSLNLHIVLRDPAVERFSRERTKCDLMDISRIYWGQHEQKLRKD 237
QY 241 QPYVMOVICOSQLEIYK--TIGSLPKALOERYLLVREDIARAPVAQTSRMVEFVGLF 298
Db 238 PHLRLREVCRSVHRIAFAATLKP-PPLRGRIYLVREDIARAPVAQTSRMVEFVGLF 296
QY 299 LPHLOTVHNITRGKMGD--HAFHTNARDALNVSOAMRWSLPEYKYSRIQKACGDAMNL 356
Db 297 TPQLEAMHNITHSGIGKRIEAFHTSSRNARNVSOAMRHALPFTKILRVQEVACAGALQ 356
QY 357 LGRHVRSQEQRNLLDLDL-----STWTYPE 383
Db 357 LGRHVRSQEQRNLLDLDL-----STWTYPE 383
RESULT 15
ABB81556
ID ABB81556 standard; Protein: 390 AA.
XX
AC ABB81556:
XX
DT 05-SEP-2002 (first entry)
XX
DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
XX
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX
XX ophthalmological.
OS Homo sapiens.
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-0927602.
XX
XX 11-AUG-2000; 2000US-325773P.
XX
XX (FUKU/) FUKUDA M N.
XX
XX (AKAM/) AKAMA T O.
XX
XX PI Fukuda MN, Akama TO;
XX
XX WPI: 2002-507643/54.
XX
XX New nucleic acid encoding corneal
XX N-acetylglucosamine-6-sulfotransferase, useful for treatment,
XX monitoring and diagnosis of macular corneal dystrophy

XX
PS Example 5; Fig 2A-B; 69pp; English.
XX
XX
CC The present invention describes human corneal
CC N-acetylglucosamine-6-sulfotransferase (1), which is able to catalyze
CC sulfation of keratan sulfate (KS). Also described is a method for
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC and detecting susceptibility to MCD. (1) is located to chromosome 16q22,
CC and has ophthalmological activity. (1) can be used to treat or prevent
CC macular corneal dystrophy types I or II. (1) makes possible treatment
CC of MCD without requiring keratoplasty or keratectomy. The present
CC sequence represents human intestinal N-acetylglucosamine-6-
CC sulfotransferase, which is given in comparison with (1) in the
CC exemplification of the present invention.
XX
SQ Sequence 390 AA:
Query Match 49.5%; Score 1008; DB 23; Length 390;
Best Local Similarity 52.3%; Pred. No. 3,7e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
QY 1 MLPRKMKLLFLVSOAILALFFHMYSHNSSL.SMKAPERMHVLYLSSWSSGSSFFVQ 60
Db 14 LLLAQTCLLFLITSRP-----GSSPAGGEDRHHVLYLSSWSSGSSFFLQ 59
QY 61 LFGQHPDVFYLMPEPAHVMWTFKOSTAMMLHNAVRLIRAVFLCDMSVFDAYMEPPRRQ 120
Db 60 LFSQHPDVFYLMPEPAHVMWTFKOSTAMMLHNAVRLIRAVFLCDMSVFDAYM-PQSRNL 118
QY 121 SSLFOWENSRALCSAPACDIIPODEIIPRAHCRLLCSQOFVEVEKACRSYSHVLYKEVR 180
Db 119 SAFENMATSRLCSPACSAFPRGTISKQDYCKTCTROPFSILARECYSYSHVLYKEVR 178
QY 181 FENLQSLYPLKDPSSLNLHIVLRDPAVERFSRERTKCDLMDISRIYWGQHEQKLRKD 240
Db 179 FENLQSLYPLKDPSSLNLHIVLRDPAVERFSRERTKCDLMDISRIYWGQHEQKLRKD 237
QY 241 QPYVMOVICOSQLEIYK--TIGSLPKALOERYLLVREDIARAPVAQTSRMVEFVGLF 298
Db 238 PHLRLREVCRSVHRIAFAATLKP-PPLRGRIYLVREDIARAPVAQTSRMVEFVGLF 296
QY 299 LPHLOTVHNITRGKMGD--HAFHTNARDALNVSOAMRWSLPEYKYSRIQKACGDAMNL 356
Db 297 TPQLEAMHNITHSGIGKRIEAFHTSSRNARNVSOAMRHALPFTKILRVQEVACAGALQ 356
QY 357 LGRHVRSQEQRNLLDLDL-----STWTYPE 383
Db 357 LGRHVRSQEQRNLLDLDL-----STWTYPE 383
Search completed: January 11, 2003, 01:18:39
Job time : 76 secs

This Page Blank (uspto)

```

FT Misc-difference 371
FT /label= Ala, Thr, Ser
FT Misc-difference 380
FT /label= Leu, Pro, Met
FT Misc-difference 382
FT /label= Gly, His, Ser
FT Misc-difference 384
FT /label= Thr, Ser, Lys
FT Misc-difference 390
FT /label= Ala, Glu
FT Misc-difference 391
FT /label= Ser, Lys
FT Misc-difference 392
FT /label= His, Glu
FT Misc-difference 394
FT /label= Arg, Glu
FT Misc-difference 395
FT /label= Asn, Ser
PN US2002061562-A1.
XX 23-MAY-2002.
XX
XX 09-AUG-2001, 2001US-0927602.
XX
XX 11-AUG-2000, 2000US-325773P.
XX
XX (FUKU/) FUKUDA M. N.
XX (AKAM/) AKAMA T O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
XX
XX New nucleic acid encoding corneal
XX N-acetylglicosamine-6-sulfotransferase, useful for treatment,
XX monitoring and diagnosis of macular corneal dystrophy -
XX
XX Example 5; Fig 2A-B; 69pp: English.
XX
XX The present invention describes human corneal
XX N-acetylglicosamine-6-sulfotransferase (I), which is able to catalyze
XX sulfation of keratan sulfate (KS). Also described is a method for
XX monitoring the effect of treatments for macular corneal dystrophy (MCD),
XX and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
XX and has ophthalmological activity. (I) can be used to treat or prevent
XX macular corneal dystrophy types I or II. (I) makes possible treatment
XX of MCD without requiring keratoplasty or keratectomy. The present
XX sequence represents a consensus N-acetylglicosamine-6-sulfotransferase
XX which is given in the exemplification of the present invention.
XX
XX Sequence 395 AA:
XX
XX Query Match 49.7%; Score 1013.5; DB 23; Length 395;
XX Best Local Similarity 55.9%; Pred. No. 9.8e-98;
XX Matches 212; Conservative 45; Mismatches 101; Indels 21; Gaps 6;
XX
XX 2 LIPKMKLLLVISOMAILLFFHMYSHNITSLSMKAPRHHVYLVSWSRSGSSEVQGL 61
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
XX 14 LLKQTKLLLFVSRP-----GPSSPAGKEXRHVYLVSWSRSGSSEVQGL 59
XX
XX 62 FGGHPDVFYILMEPAWHVMTFKOSTAMLMKAVDILRAVFLCDMSVFDAYMEGRPRQS 121
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
XX 60 FSGHPDVFYILMEPAWHVMTLQSSAATLHMAVADLKRSVFLCDMDVFDAYL-PWRRLNS 118
XX
XX 122 SLFQWMSRALCAPADIIIPODEIIPRAHORLICSOQFEVEVERKACRSYSHVLEKVF 181
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
XX 119 DLFQWMSRALCAPACAPAFKRXISEYCKPLCARQPPXILAREACRSYSHVLEKVF 178
XX
XX 182 FNLQSLVPLKDPCLNLIHYLVDPRAVPSRRTKGDMLIDSRIVYMGQHEOKLKKEQD 241
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
XX 179 FNLQSLVPLKDPCLNLIHYLVDPRAVPSRRTKGDMLIDSRIVYMGQHEOKLKKEQD 237
XX

```

```

QY 242 PYVWQVICOSOLEIYK--TIQSLPKALQERILVRYEDLARAAYAGTSRMTEFEVGL 299
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 238 XLRVYREVCRSHVRIAEATLKP--PPFLGRVRLVRFEDLAREPLAEIRAYATGKXLT 296
XX
QY 300 PHLQFWNHTITGRKGMG--DHAEHTNARDALNVSQAWMSLPEYEVSPLOKACDAMNLL 357
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 297 POLEAMINHTITGSGPGARREAFKTSRNALNVSQAWMALPFAKIRVQGLCAGALQL 356
XX
QY 358 GYRHSVSEDEQRNLLDEL 376
XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 357 GYRHSVSEDEQRNLLDLV 375
XX
RESULT 14
AAV72639
ID AAV72639 standard; Protein: 390 AA.
XX
XX AAV72639;
XX
XX 02-MAY-2001 (first entry)
XX
XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).
XX
XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
XX therapy; selectin binding inhibitor; gene therapy; inflammation;
XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX polyarteritis nodosa; polypositis; systemic sclerosis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1.
XX
XX Homo sapiens.
XX
XX WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000MO-US19741.
XX
XX 20-JUL-1999; 99US-0144694.
XX
XX 13-JUL-2000; 2000US-0593828.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
XX
XX WPI; 2001-138471/14.
XX
XX N-PSDB; AAD02697, AAD02698, AAD02699.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications -
XX
XX Claim 3; Fig 1; 128pp: English.
XX
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX alpha). GST-4 gene is found on chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX selectin with a non-sulphated selectin ligand, GST is also useful
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation,
XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX nodosa, polypositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX

```

QY 298 FLPHLOTWVHNTTRGKMG--DHAFTNARDALNVQAMRWSLPEYKVSRLQKAGDAMN 355
 Db 295 LTPOLQTIHNTHTGSGPGARREAFKTSRDLVSQAMRHLPFAKIRVQELCGALQ 354
 QY 356 LGGYRVHSEEOQRNLLIDL 376
 Db 355 LGGYRVHSELEQRDLIDL 375

RESULT 12

AAU11275
 ID AAU11275 standard; Protein: 395 AA.

AC AAU11275:

DT 12-MAR-2002 (first entry)

XX Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.

XX Mouse: beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuk;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.

XX Mus musculus.

XX WC20185177-A1.

XX 15-NOV-2001.

XX 10-MAY-2001; 2001MO-US15452.

XX 11-MAY-2000; 2000MO-0569320.

XX (BURN-) BURNHAM INST.

XX Fukuda M, Yeh J, Hiraoka N.

XX WPI; 2002-075226/10.

XX N-PSDB; AAS16948.

PT New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 PT intestinal GlcNAc 6-sulfotransferase

PS Claim 28; Fig 10; 98pp; English.

XX The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active
 CC fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The
 CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC beta1,3GNT that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds beta1,3GNT,
 CC and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of beta1,3GNT. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents mouse I-GlcNAc6ST.

XX Sequence 395 AA;

Query Match 49.9%; Score 1017.5; DB 23; Length 395;
 Best Local Similarity 56.4%; Pred. No. 3.7e-98;
 Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

QY 1 MLPK--KKALLFLVSQAIALLFFHMTSHNLSLMSKAQEBRMAYLVLSMRGSSSFV 58
 Db 1 MRLPRFSSTVMISLMVQGIIVF--LVSQVPS-SPAGLGEYHVLVLSMRGSSSFV 56
 QY 59 GOLFGQHPDVEFYLMDEPAMVHMTFKOSTAMLMMAVRDLIRAFVLDMSVFAVMEPGPR 118
 Db 57 GOLFSQHPDVFYIMDEPAMVHMTDLSSGSAAPALHMAVRDLIRVFLCDMVFAYIL-PMNR 115
 QY 119 ROSSLFQWENSRLALCSAPACDIIPQDEILPRAHCRILCSQPFVEYKACRSYSHVLE 178
 Db 116 NISDLFQWAVSRALCSPPVCEAFARGNISSEYVCKPLCATRFFGIAQECSSYSHVLE 175
 QY 179 VRFNFQSLSYPLIKDLSLMLHIVLYRDRAYRSEPRFKGDMIDSRVVMQCHQKTK 238
 Db 176 VRFNFQVLYPLSDPALMLRYHLVDRPRAVLRSEQYAKALARDNGIVLTNGTW-E 234
 QY 239 EDQPYVMQVICOQSOLEIK-TIQSLPKALOERYLLVRYEDLARAPVQTSRMYEFGIE 297
 Db 235 ADBRLVNVNEGVCRSHVRIAEALHKPPPLQDRVLRVYEDLARQPLVYIRELYAFTGIG 294
 QY 298 FLPHLOTWVHNTTRGKMG--DHAFTNARDALNVQAMRWSLPEYKVSRLQKAGDAMN 355
 Db 295 LTPOLQTIHNTHTGSGPGARREAFKTSRDLVSQAMRHLPFAKIRVQELCGALQ 354
 QY 356 LGGYRVHSEEOQRNLLIDL 376
 Db 355 LGGYRVHSELEQRDLIDL 375

RESULT 13

ABB81555
 ID ABB81555 standard; Protein: 395 AA.

XX ABB81555;

XX 05-SEP-2002 (first entry)

XX Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO.3.

KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GlcNAc6ST;
 KW cornual; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.

XX Homo sapiens.

XX Mus musculus.

XX Synthetic.

FT Key Location/Qualifiers
 FT Misc-difference 10 /Label= Ala, Thr, Val
 FT Misc-difference 13 /Label= Ala, Val, Ser
 FT Misc-difference 20 /Label= Phe, Cys, Gly
 FT Misc-difference 39 /Label= Ala, Asp, Glu
 FT Misc-difference 96 /Label= Val, Met, Ile
 FT Misc-difference 142 /Label= Ala, Thr, Asn
 FT Misc-difference 147 /Label= Ala, Asp, Glu
 FT Misc-difference 159 /Label= Thr, Ser, Gly
 FT Misc-difference 238 /Label= Gly, His, Arg
 FT Misc-difference 294 /Label= Ser, Thr, Gly

CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,
CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
CC infections; cell proliferative disorders such as actinic keratosis,
CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
CC epilepsy; endocrine disorders such as disorders of the hypothalamus
CC and pituitary resulting from lesions such as primary brain tumours,
CC adenomas, infection associated with pregnancy, aneurysms, vascular
CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
CC glaucoma, pigmentosa, metabolic disorders such as Addison's disease,
CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
CC disorders. The present sequence is human DME-5 protein.
XX
XX
SQ Sequence 395 AA:
Query Match 50.0%; Score 1019; DB 23; Length 395;
Best Local Similarity 54.5%; Pred. No. 2.6e-98;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;
QY 2 LIPKMKLLFTVSQMAILALFFHMYSHNITSLSMKAOBERNHVILSSWRSGSSFFVGL 61
DB 14 LLAQTFLLFLFVSRRP-----GPSSPAGCBARVAVLTLSSWRSGSSFFVGL 59
QY 62 FQGHDPVFTLMEPAHVMWTFKOSTAMKLMAYRDLIRAVFLCDMSVDPAYMEPPGR 121
DB 60 FQGHDPVFTLMEPAHVMWTFKOSTAMKLMAYRDLIRAVFLCDMSVDPAYL-PWRNIN 118
QY 122 SLFQWNSRALCSAPACDIIPODEIIPRAHCRLLCSQGFVEVEKACRSYSHVLEKVF 181
DB 119 DLFQWAVSRALCSPPACSAFPRGAISSAVCKFLCAROSTFLAREACRSYSHVLEKVF 178
QY 182 FNIQSLYPLKDPSLNLTIVLVRDPRAVRSRERTKGLMDSRTVMQHEKLLKDEQ 241
DB 179 FNIQSLYPLKDPALNLTIVLVRDPRAVRSRERTKAKALRANGVLVNGTWV-EADP 237
QY 242 PYYVMQVTCOSQLEIK--TIQSLPRALQERYLLVREDIARAPVAQTSRMYEFGLE 299
DB 238 GLRVREVCSHVRIAEATLKP--PPLRGRYRLVREDIARAPLAEIRALVFTGLSLT 296
QY 300 PHLQTVNHTTRKGMG--DHAFTNARQALNLSQAMRNSLPREKVSRLQKAGCDAMNL 357
DB 297 POEAMVHNHTHSGPARREAFKTSRNALNLSQAMRHALLPFAKIRVQELCAGALL 356
QY 358 GYRHVSEQEQRLDLDL-----STWTVPRQIH 386
DB 357 GYRPVYSEDEQRNALDLVLRGLNGTWTASSTASH 392
RESULT 11
AAV72638
ID AAV72638 standard; Protein; 395 AA.
XX
XX AAV72638;
XX
XX 02-MAY-2001 (first entry)
XX
XX Mouse glycosyl sulfotransferase-4 (GST-4).
DE Mouse glycosyl sulfotransferase-4 (GST-4).
XX
XX Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
KM therapy; selectin binding inhibitor; gene therapy; inflammation;
KM systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KM polyarteritis nodosa; polyvasculitis; systemic sclerosis; dermatitis;
KM glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;

KM Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KM demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KM myocaditis; adult respiratory distress syndrome; eczema; psoriasis;
KM asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 8e1.
XX
XX
OS Mus musculus.
XX
XX W0200106015-A1.
XX
XX 25-JAN-2001.
PD
XX
XX 19-JUL-2000; 2000WD-US19741.
PF
XX
XX 20-JUL-1999; 99US-0144694.
PR
XX 13-JUL-2000; 2000US-0533828.
XX
XX (REGC) UNIV CALIFORNIA.
PA
XX
XX Rosen SD, Lee JK, Hemmerich S;
PI
XX WPL: 2001-138471/14.
DR
XX N-PSDB; AAD02696.
PT
XX
PS New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications -
Claim 3; Fig 2; 128pp; English.
XX
XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
CC gene is found on chromosome 8e1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, peritonitis
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
XX
XX
SQ Sequence 395 AA:
Query Match 49.9%; Score 1017.5; DB 22; Length 395;
Best Local Similarity 56.4%; Pred. No. 3.7e-98;
Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;
QY 1 MLPRK--KKKLLFTVSQMAILALFFHMYSHNITSLSMKAOBERNHVILSSWRSGSSFFV 58
DB 1 MLPRFSSIVMLSLMQTGILVF--LVSRQVPS-SPAGLGERNHVILSSWRSGSSFFV 56
QY 59 GOLFGHDPVFTLMEPAHVMWTFKOSTAMKLMAYRDLIRAVFLCDMSVDPAYMEPPGR 118
DB 57 GOLFGHDPVFTLMEPAHVMWTFKOSTAMKLMAYRDLIRAVFLCDMSVDPAYL-PWR 115
QY 119 ROSLFFQWNSRALCSAPACDIIPODEIIPRAHCRLLCSQGFVEVEKACRSYSHVLEK 178
DB 116 NISDLFQWAVSRALCSPPACSAFPRGAISSAVCKFLCAROSTFLAREACRSYSHVLE 175
QY 179 VRFENLQSLYPLKDPSLNLTIVLVRDPRAVRSRERTKGLMDSRTVMQHEQDKLK 238
DB 176 VRFENLQSLYPLKDPALNLTIVLVRDPRAVRSRERTKAKALRANGVLVNGTWV-E 234
QY 239 EQQPYVMQVTCOSQLEIK--TIQSLPRALQERYLLVREDIARAPVAQTSRMYEFGLE 297
DB 235 ADPRILRVAVNCSHVRIAEALHKPPPLQDRIVLVRREDIAROPLVYIRALVFTGLG 294

```

XX AC ABB81554;
XX DT 05-SEP-2002 (first entry)
XX DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
XX KW Corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX KW Ophthalmological; chromosome 16q22.
XX OS Homo sapiens.
XX PN US2002061562-A1.
XX PD 23-MAY-2002.
XX PF 09-AUG-2001; 2001US-0927602.
XX PR 11-AUG-2000; 2000US-325773P.
XX PA (FUKU/) FUKUDA M N.
XX PA (AKAM/) AKAMA T O.
XX PI Fukuda MN, Akama TO;
XX DR WPI: 2002-507643/54.
XX DR N-PSDB: ABB89506.
XX PT New nucleic acid encoding corneal
XX PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
XX PT monitoring and diagnosis of macular corneal dystrophy -
XX PS Claim 13; Fig 1A-D; 69pp; English.

CC The present sequence represents human corneal
CC N-acetylglucosamine-6-sulfotransferase (1), which is able to catalyze
CC sulfation of keratan sulfate (KS). Also described is a method for
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC and detecting susceptibility to MCD. (1) is located to chromosome 16q22,
CC and has ophthalmological activity. (1) can be used to treat or prevent
CC macular corneal dystrophy types I or II. (1) makes possible treatment
CC of MCD without requiring keratoplasty or keratectomy.
XX CC
XX SQ Sequence 395 AA;

Query Match 50.0%; Score 1019; DB 23; Length 395;
Best Local Similarity 54.5%; Pred. No. 2,6e+98;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LERKMKLLFLVSOAIIALPFFHMYSHNITSLMKRAOBERMHVLYLSMRSGSSFVQGL 61
DB 14 LLLAQFTLLFLVSRP-----GPSPPAGGEARVHVLYLSMRSGSSFVQGL 59
QY 62 FGOHPDYFYIMEBAKHYWMTFKQSTAMMLHMAVRLIRAVFLCDMSVFPAYMEPPGRROS 121
DB 60 FNOHPDYFYIMEBAKHYWMTLLSQGSATLHMAVRLIRAVFLCDMDVFPAYL-PMRRNLS 118
QY 122 SLFQWENSRAICSPACDIIPODEIIPRAHCRLLCSQPFEEVEKACRSYSHVLEKVEF 181
DB 119 DLFQWANSRAICSPACSAFPRGAISSEAVCKPLCAROSTTLARECRSHVLEKVEF 178
QY 182 FNIQSLPLKDPSSLNLIHVLRDPRAVFRSREKRGKGLMDISRIVMGOHEOKLEKEDO 241
DB 179 FNIQSLPLKDPSSLNLIHVLRDPRAVFRSREKRGKGLMDISRIVMGOHEOKLEKEDO 241
QY 242 PYYVMVOICSOLEIK--TIOSLPKALOERYTLVREYEDLARPVQOTSRMVEFVLEEL 299
DB 238 GLRVRFVCSHVRIAEATLKP-PPPLKORYKLVRFEDARPELAEIRALIAFTGLSLT 296
QY 300 PHLQIWNHTTRKSGM--DHAFHTNARDALNVSQAMRMSLPYEKYSRLQKAGDAMNLL 357
DB 297 PDLQIWNHTTRKSGM--DHAFHTNARDALNVSQAMRMSLPYEKYSRLQKAGDAMNLL 356

```

```

QY 356 GYRHWSEDEQRNLLDLL-----STWVPEQHT 396
DB 357 GYRHWSEDEQRNLLDLLPRLGNGFTWASSTASH 392

RESULT 10
AAE15438
ID AAE15438 standard; Protein; 395 AA.
XX AAE15438;
AC AAE15438;
XX 12-MAR-2002 (first entry)
DE Human drug metabolizing enzyme (DME)-5.
XX Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
XX inflammatory disorder; acquired immune deficiency syndrome; infection;
XX AIDS; adult respiratory distress syndrome; cell proliferative disorder;
XX allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
XX cancer; endocrine disorder; hypohalimus disorder; pituitary disorder;
XX gastrointestinal disorder; metabolic disorder; developmental disorder;
XX liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
XX goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
XX DME-5.
XX OS Homo sapiens.
XX FH Key
XX FH Peptide
XX FT /label= Signal_peptide
XX FT Protein
XX FT /note= "Human mature DME-5 protein"
XX FT Peptide
XX FT /label= Signal_peptide
XX FT Protein
XX PD /note= "Human mature DME-5 protein"
XX WO200179468-A2.
XX 25-OCT-2001.
XX 12-APR-2001; 2001WO-US11869.
XX 13-APR-2000; 2000US-197590P.
XX 19-APR-2000; 2000US-198403P.
XX 28-APR-2000; 2000US-200185P.
XX 05-MAY-2000; 2000US-202234P.
XX 11-MAY-2000; 2000US-203509P.
XX (INCY-) INCYTE GENOMICS INC.
XX Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
XX Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjapala MS;
XX Ganchi AR, Reddy R, Khan FA, Baughn WR, Ramkumar J, Griffin JA;
XX Au-Young J;
XX WPI: 2002-066363/09.
XX N-PSDB: AAD24670.
XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,
XX useful for diagnosing, treating, or preventing disorders associated
XX with aberrant expression of DME such as allergy, anemia, asthma,
XX infertility -
XX Claim 1a; Page 131-132; 143pp; English.

CC The invention relates to human drug metabolizing enzymes referred as
CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
CC of the invention are useful for assessing toxicity of test compounds
CC and in gene therapy. Sequences of the invention are useful in the
CC diagnosis, prevention and treatment of autoimmune/inflammatory
CC disorders such as acquired immune deficiency syndrome (AIDS), adult

```


RESULT 6
 AAB41947
 ID AAB41947 standard; protein; 418 AA.
 AC AAB41947;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnereary; antipsoptic; antiparkinsonian; noctropic; neuroprotective;
 KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;
 KW antineurotic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 PN WO200058473-A2.
 PD 05-OCT-2000.
 PF 31-MAR-2000; 2000WO-US08621.
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 PA (CURA-) CURAGEN CORP.
 PI Shinkets RA, Leach M;
 DR WPI: 2000-602362/57.
 DR N-PSDB: AAC76156.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PS Claim 11: Page 2599-2600; 5507pp; English.
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnereary;
 CC antipsoptic; antiparkinsonian; noctropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antifungal; antineurotic;
 CC antihypertensive; antineurotic; antineoplastic; antineurotic;
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SEQ Sequence 418 AA:
 Query Match 59.2%; Score 1205.5; DB 21; Length 418;
 Best Local Similarity 61.2%; Pred. No. 6.7e-118;
 Matches 241; Conservative 47; Mismatches 79; Indels 27; Gaps 6;
 QY 1 MLPPKRLLLFLVSOMATLALFFHYSNHISSLSMKAPERMHVLVLSWRSSSEVQ 60
 DB 41 LLIAQTCLLLFTISRP-----GSSPAGEDRVHVLVLSWRSSSEVQ 86
 QY 61 LFGQHGVYVYVLEPRANHYMTFKQSTAMMLHMAVBDLIRAVFLCMSVPDAMERGPRQ 120
 DB 87 LFSQHPGVYVLEPRANHYMTFKQSTAMMLHMAVBDLIRAVFLCMSVPDAMERGPRQ 146
 QY 121 SSLFQWENSRALCSAPACDIIPQDEITPRACRLLCSQPFVEVAKRSYSHVYLKEVR 180
 DB 147 SSLFQWENSRALCSAPACDIIPQDEITPRACRLLCSQPFVEVAKRSYSHVYLKEVR 206
 QY 181 FFLQSLYFLLDPSLNLIHVLVPRPRAVFSRERTKGLMIDRIYVGGHQLKED 240
 DB 207 FFLQSLYFLLDPSLNLIHVLVPRPRAVFSRERTKGLMIDRIYVGGHQLKED 265
 QY 241 QPYVWQVICSQLEIYK--TTQSLPKALQERYLLVREYEDLARAPVQTSRMYEFVGLF 298
 DB 266 PHLRITREYCRSHVRIAEATLKP--PPLRGYRLVREFDLARPELAEIRALYFTGLL 324
 QY 299 LPHLQVWVNIIRGKQMD--HAFTNARDALVNSQAMRSLPYEKVSLQACGDAMNL 356
 DB 325 TPLEAMINHIHIGSGICKEAFHTSSRNARNVSOAWRHALPFTKILRVQECAGALQ 384
 QY 357 LGRHYRSFQEOBNLLDL-----STWVPE 383
 DB 385 LGRPYVSADQQRDLIDLPRGPDHFSWASPD 418
 RESULT 7
 ABB81557
 ID ABB81557 standard; protein; 418 AA.
 AC ABB81557;
 PD 05-SEP-2002 (first entry)
 DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
 KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; G1CNA65T;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 OS Mus musculus.
 PN US2002061562-A1.
 PD 23-MAY-2002.
 PF 09-AUG-2001; 2001US-0927602.
 PR 11-AUG-2000; 2000US-325773P.
 PA (FUKU/) FUKUDA M N.
 PA (AKAW/) AKAWA T O.
 PI Fukuda MN, Akawa TO;
 DR WPI: 2002-507643/54.
 XX
 PT New nucleic acid encoding corneal
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
 PT monitoring and diagnosis of macular corneal dystrophy -
 PS Example 5; Page 24-25; 69pp; English.
 CC The present invention describes human corneal

XX ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; anti-leer;
 KW anti-inflammatory; antipsoriatic; antidiabetic; dermatological;
 KW anti-allergic.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO2001085177-A1.
 PN
 XX
 XX 15-NOV-2001.
 PD
 XX
 XX 10-MAY-2001; 2001WO-US15452.
 PF
 XX
 XX 11-MAY-2000; 2000US-0569320.
 PR
 XX
 XX (BURN-) BURHAM INST.
 PA
 XX
 XX Fukuda M, Yeh J, Hiraoka N;
 PI
 XX
 XX WPI: 2002-075226/10.
 DR
 XX
 XX N-PSDB: AAS16947.
 PT
 XX
 XX New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 PT intestinal GlcNAc 6-sulfotransferase
 XX
 PS Claim 21: Fig 4; 98pp; English.
 CC
 XX
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC betal,3-N-acetylglucosaminyltransferase (betal,3gnt) or an active
 CC fragment, where betal,3gnt directs expression of a MECA-79 antigen. The
 CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC betal,3gnt that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds betal,3gnt,
 CC and/or a betal,3gnt antisense nucleic acid molecule, L-selectin
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of betal,3gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents human LST-2.
 CC

| SQ | Sequence | 380 AA: |
|----|---|---------------------------------------|
| | Query Match | 98.5%; Score 2008; DB 23; Length 380; |
| | Best Local Similarity | 100.0%; PId, No. 2, 1e-202; |
| | Matches 380; Conservative | 0; Mismatches 0; Indels 0; Gaps 0. |
| QY | 7 MKLLFLVSMATLALFFHMYSHNISTSLSKAQPFMHVLVLSWRSGSSFFQLGQHP | 66 |
| Dd | 1 MKLLFLVSOMATLALFFHMYSHNISTSLSKAQPERMHLVLSWRSGSSFVQLGQH | 60 |
| QY | 67 DVFILMEPAHHVMYTKOSTANMLLMANRDLIRAVFLCDMSVPAYIHPPGRQSSLFQM | 126 |
| Dd | 61 DVFLMEPAHHVMYTKOSTANMLLMANVRDLIRAVFLCDMSVEPAYHEPGRQSSLFQM | 120 |
| QY | 127 ENSRALCSAPACDIIPDEIIPRAHCRIKLCQQPEVEVERKASRSYHAVLKEVREFNLQS | 186 |
| Dd | 121 ENSRALCSAPACDIIPDEIIPRAHCRIKLCQQPEVEVERKASRSYHAVLKEVREFNLQS | 180 |
| QY | 187 LYPILLDPSLNLIHYLVADPRAYFRSRETRKGLMDISRYMQGHQOKLKREDQPYIVA | 246 |
| Dd | 181 LYPILLDPSLNLIHYLVADPRAYFRSRETRKGLMDISRYMQGHQOKLKREDQPYIVA | 240 |

| | | | |
|----|-----|---|-----|
| QY | 247 | QVTCOSQJTEIKTQJOSJLTKALOBERLLVRYADDLARAVAOJSRMYEVSGLPFIPLHQTW | 306 |
| Db | 241 | QVTCOSQJTEIKTQJOSJLTKALOBERLLVRYADDLARAVAOJSRMYEVSGLPFIPLHQTW | 300 |
| QY | 307 | HNIFPGKMGHGHHTNARNRDALNYSOAMRWLSLPYEXYSJLOKACGDAMNLLGVRHRSQ | 366 |
| Db | 301 | HNIFPGKMGHGHHTNARNRDALNYSOAMRWLSLPYEXYSJLOKACGDAMNLLGVRHRSQ | 360 |
| QY | 367 | FORNLLDPLSTWVPEQIH | 386 |
| Db | 361 | FORNLLDPLSTWVPEQIH | 380 |

RESULT 4

ID AAY79219 standard; Protein; 386 AA

AC AAY79219;

DT 19-JUN-2000 (first entry)

Human transferase TRNSFS-11.

KW Transferrase; TRNSFS-11; human; antitumour; cell proliferation;

KW genetic disorder; neurological disorder; reproductive disorder;
 KW smooth muscle; immunological disorder; inflammation

KW diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.
XX

OS Homo sapiens.
XY

| EH | key | Location/qualifiers |
|----|---------------|---------------------|
| ET | Modified-site | 121 |

| | | | |
|-----|---------------|-----|------------------------------------|
| FTT | Modified-site | 107 | /note="potential O-phosphorylation |
|-----|---------------|-----|------------------------------------|

| | | | |
|----|---------------|-----|------------------------------------|
| ET | Modified-site | 217 | /note="potential O-phosphorylation |
|----|---------------|-----|------------------------------------|

Modified-site 252

| | | | |
|----|---------------|-----|------------------------------------|
| FT | Modified-site | 364 | /note= potential 0-phosphorylation |
|----|---------------|-----|------------------------------------|

| | | | |
|-----|---------------|--------|-----------------------------|
| 4.1 | Modified-site | /note- | potential o-phosphorylation |
| ET | | 380 | |

| | | | | |
|----|---------------|----|-------|------------------------------|
| FT | Modified-site | 35 | /nuc- | potential to phosphorylation |
|----|---------------|----|-------|------------------------------|

| FT | Modified-site | 50 | /nuc- | potential | to phosphorylation |
|----|---------------|----|-------|-----------|--------------------|
| 11 | | | | | |

| FT | Modified-site | 7 nucleotides | Potential of phosphorylation |
|----|---------------|---------------|------------------------------|
| | 81 | | |

| | | | | |
|----|---------------|-----|--------|-----------------|
| FT | Modified-site | 287 | 7.400e | Phosphorylation |
|----|---------------|-----|--------|-----------------|

| | | | |
|----|---------------|-----|-----------------|
| FT | Modified-site | 243 | phosphorylation |
|----|---------------|-----|-----------------|

FT Modified-site 30

| | | | |
|----|---------------|-----|-------|
| FT | Modified-site | 308 | |
|----|---------------|-----|-------|

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----|---------------|-----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|
| FT | Modified-site | 329 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |
|----|---------------|-----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|

| FT | Domain | 7.23 |
|----|--------|------|
| | | |

XX

XX

XX

XX

PR 04-NOV-1998; 98US-0186779.

XX

PT Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 PS Claim 2; Fig 1; 59pp; English.

CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.

XX Sequence 386 AA:

Query Match 100.0%; Score 2038; DB 20; Length 386;
 Best Local Similarity 100.0%; Pred. No. 1.5e-205;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSQMAIIALFFHMYSHNITSLSKAKQPERMHVLYLSMRSQSSFFVQ 60
 DB 1 MLPPKMKLLFLVSQMAIIALFFHMYSHNITSLSKAKQPERMHVLYLSMRSQSSFFVQ 60
 QY 61 LFGQHPDVFYLMPPAHVWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 DB 61 LFGQHPDVFYLMPPAHVWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 QY 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVLEKVR 180
 DB 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVLEKVR 180
 QY 181 FENLQSLVPLKDPSPNLHIVLVRDPRAVFRSRETKGDMIDSRIVMGQHEQKLKED 240
 DB 181 FENLQSLVPLKDPSPNLHIVLVRDPRAVFRSRETKGDMIDSRIVMGQHEQKLKED 240
 QY 241 QPYVWQVIOCSOLEIKTIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFVGLFLP 300
 DB 241 QPYVWQVIOCSOLEIKTIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFVGLFLP 300
 QY 301 HLOTWVHNITRKGMDGDAFTNARDALNVSQAWMSLPEYKVSRLQKACGDAMNLGVR 360
 DB 301 HLOTWVHNITRKGMDGDAFTNARDALNVSQAWMSLPEYKVSRLQKACGDAMNLGVR 360
 QY 361 HVSEDEQRNLLDLSTWVPEQIH 386
 DB 361 HVSEDEQRNLLDLSTWVPEQIH 386

RESULT 2
 ID AAM93309 standard; Protein; 386 AA.

XX AAM93309;

DT 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 2817.

KM Human; full length cDNA; cDNA synthesis; oligo-capping.

OS Homo sapiens.

PN EP1130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

PA (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR WPI: 2001-524255/58.
 DR N-PDB: AAK94229.

PT 830 primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS Claim 8; SEQ ID NO 2817; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 386 AA:

Query Match 99.9%; Score 2035; DB 22; Length 386;
 Best Local Similarity 99.7%; Pred. No. 3.1e-205;
 Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSQMAIIALFFHMYSHNITSLSKAKQPERMHVLYLSMRSQSSFFVQ 60
 DB 1 MLPPKMKLLFLVSQMAIIALFFHMYSHNITSLSKAKQPERMHVLYLSMRSQSSFFVQ 60
 QY 61 LFGQHPDVFYLMPPAHVWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 DB 61 LFGQHPDVFYLMPPAHVWMTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEGPRRQ 120
 QY 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVLEKVR 180
 DB 121 SSLFQWENSRALCSAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVLEKVR 180
 QY 181 FENLQSLVPLKDPSPNLHIVLVRDPRAVFRSRETKGDMIDSRIVMGQHEQKLKED 240
 DB 181 FENLQSLVPLKDPSPNLHIVLVRDPRAVFRSRETKGDMIDSRIVMGQHEQKLKED 240
 QY 241 QPYVWQVIOCSOLEIKTIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFVGLFLP 300
 DB 241 QPYVWQVIOCSOLEIKTIQSLPKALQERYLLVRYEDLARAPVQTSRMVEFVGLFLP 300
 QY 301 HLOTWVHNITRKGMDGDAFTNARDALNVSQAWMSLPEYKVSRLQKACGDAMNLGVR 360
 DB 301 HLOTWVHNITRKGMDGDAFTNARDALNVSQAWMSLPEYKVSRLQKACGDAMNLGVR 360
 QY 361 HVSEDEQRNLLDLSTWVPEQIH 386
 DB 361 HVSEDEQRNLLDLSTWVPEQIH 386

RESULT 3
 ID AAO11274 standard; Protein; 380 AA.

XX AAO11274;

DT 12-MAR-2002 (first entry)

DE Human L-selectin sulfotransferase-2 (LSST-2) protein.

KM Human; beta1,3Gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 L-selectin; L-selectin sulfotransferase-2; Cronin's disease; diabetes;

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2003, 21:45:02 ; Search time 70 Seconds

(without alignments)
734.782 Million cell updates/sec

Title: US-09-816-825-2

Perfect score: 2038
Sequence: 1 MLPLPKMKLLFLVVSOMATL.....EQRNLLDLSTWVPEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|--------|--------------------|-----|--------------------------------|
| 1 | 2038 | 100.0 | 386 | 20 AAY39918 Human glycosyl sul |
| 2 | 2035 | 99.9 | 386 | 22 AAM93309 Human polypeptide |
| 3 | 2008 | 98.5 | 380 | 23 AAU11274 Human l-selectin s |
| 4 | 1936 | 95.0 | 386 | 21 AAY92919 Human transferrase |
| 5 | 1513.5 | 74.3 | 388 | 20 AAY39919 Human glycosyl sul |
| 6 | 1203.5 | 59.2 | 418 | 21 AAB41947 Human ORFX ORF1711 |
| 7 | 1028 | 50.4 | 418 | 22 AAB81557 Mouse intestinal N |
| 8 | 1019 | 50.0 | 395 | 22 AAY72640 Human glycosyl sul |
| 9 | 1019 | 50.0 | 395 | 23 ABB81554 Human corneal N-ac |
| 10 | 1019 | 50.0 | 395 | 23 AAE15438 Human drug metabol |

| | | | | |
|----|--------|------|------|-------------|
| 11 | 1017.5 | 49.9 | 395 | 22 AAY72638 |
| 12 | 1017.5 | 49.9 | 395 | 23 AAU11275 |
| 13 | 1013.5 | 49.7 | 395 | 23 ABB81555 |
| 14 | 1008 | 49.5 | 390 | 22 AAY72639 |
| 15 | 1008 | 49.5 | 390 | 23 ABB81556 |
| 16 | 821 | 40.3 | 171 | 23 ABB81560 |
| 17 | 598.5 | 29.4 | 483 | 20 AAY31656 |
| 18 | 591.5 | 29.0 | 530 | 22 AAB95367 |
| 19 | 587.5 | 28.8 | 484 | 20 AAY31657 |
| 20 | 587.5 | 28.8 | 531 | 23 AA069414 |
| 21 | 549 | 26.9 | 411 | 19 AAM61100 |
| 22 | 519.5 | 25.5 | 169 | 23 ABB81558 |
| 23 | 517.5 | 25.1 | 169 | 23 ABB81559 |
| 24 | 500.5 | 23.7 | 458 | 18 AAW06480 |
| 25 | 482 | 23.6 | 479 | 19 AAW02863 |
| 26 | 336 | 16.5 | 169 | 23 ABB81561 |
| 27 | 335.5 | 16.5 | 179 | 23 ABB81562 |
| 28 | 293 | 14.4 | 174 | 23 ABB81563 |
| 29 | 210 | 10.3 | 363 | 22 ABB64512 |
| 30 | 186.5 | 9.2 | 183 | 22 ABB68582 |
| 31 | 109 | 5.3 | 1207 | 22 AAY72643 |
| 32 | 108.5 | 5.3 | 596 | 22 AAY72641 |
| 33 | 108.5 | 5.3 | 1222 | 22 AAY72642 |
| 34 | 100.5 | 4.9 | 775 | 23 ABB97942 |
| 35 | 96.5 | 4.7 | 791 | 23 AAE22917 |
| 36 | 94 | 4.6 | 3588 | 14 AAR34712 |
| 37 | 92.5 | 4.5 | 2354 | 22 ABB60511 |
| 38 | 92 | 4.5 | 359 | 20 AAY06628 |
| 39 | 92 | 4.5 | 359 | 21 AAY84309 |
| 40 | 91.5 | 4.5 | 5373 | 22 AAU14603 |
| 41 | 91.5 | 4.5 | 1847 | 21 AAY52002 |
| 42 | 91 | 4.5 | 1847 | 21 AAY51631 |
| 43 | 89.5 | 4.4 | 346 | 23 ABB81056 |
| 44 | 89 | 4.4 | 566 | 22 ABB30335 |
| 45 | 89 | 4.4 | 920 | 23 ABB25932 |

ALIGNMENTS

RESULT 1
AAY39918
ID AAY39918 standard; Protein: 386 AA.
XX
AC AAY39918;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human glycosyl sulfotransferase-3 protein sequence.
XX
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; human;
KW secondary lymph organ.
XX
OS Homo sapiens.
XX
PN WO949018-A1.
XX
PD 30-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
PA (REGC) UNIV CALIFORNIA.
PA (SYNT) SYNTAX USA INC.
XX
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX WPI, 1999-580442/49.
XX DR N-PSDB; AA20792.
XX

Db 105 TTGCTGCGGCGCGCGCTTCACCTCGTCCATCATGTT 64

Search completed: January 10, 2003, 17:55:18
Job time : 296 secs

RESULT 15
US-09-815-242-7888

Sequence 7888, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

Prokaryotes

FILE REFERENCE: ELITRA.01A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 1410

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 7888

LENGTH: 2838

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(2838)

US-09-815-242-7888

Query Match 2.0%; Score 40; DB 10; Length 2838;

Best Local Similarity 44.8%; Pred. No. 0.31;

Matches 154; Conservative 0; Mismatches 190; Indels 0; Caps 0;

QY 634 CTGACAGTCTCTGTGACGTACAGCCCTTTGAGTGTGAGAGGCCCTGCCCTCTTA 693

Db 345 CGGTACCCCGGCTGCGCGGACGACATCCGCTGAGGCGACGCGTACGCCAGAT 404

QY 694 CAGCCAGTGTGCTCAAGAGAGTGCCTTCTCACTGCTAGTCCCTTACCCGCTGCT 753

Db 405 GGTGACAGAGTCTCTGCGGCTGCCGAGAGGACAGCATGCTGCTGCGGCGGCTGAT 464

QY 754 GAAGAGCCCTCTCTCAACCTGCTATGCTGACCTGTCCGGGAGCCCGGGCGCTGTT 813

Db 465 CCGGAGGCGCAAGGCGGACCTGCGCGGTGAGACAGATGCGCGGCGGAGGCTTGT 524

QY 814 CCGTCCGAGAGACGACAAAGGAGATCTCATGATGACAGTGCATGTTGATGGGCA 873

Db 525 CCGGCGCGGCTGACGCGGAGCTCTACGAGCTGACGAGAGTCCGGAAGTGAAGCA 584

QY 874 GCATGACGAGAACTCAAGAGAGAGACCAACCTACTATGTATGACAGTCACTTCCA 933

Db 585 GAAGAGACGACGATCGATGTGTGTGAGACCGCTTCAAGTTGCGCGGACCTCCAGCA 644

QY 934 AAGCGAGTGAAGATCTACAGAGACCATCCATCTTGGCCAGG 977

Db 645 ACGCTGCGGAGTGTGAGACCGCCCTGTCTCTGCGGCGAG 688

This Page Blank (uspto)

| | | | | |
|---|-----|--|-----|--|
| QY | 708 | TCAAAGAGAGTGCCTTTCTTCAACTGCAGTGCCTTTACCCGCTGCTGAAGACCCCTCC | 767 | APPLICANT: Fan, Liqun |
| Db | 3 | TAAAGGGTGTGGGCTTTTCAGCTGGGGGCTTTGGCCCACTGGTGGAGACCCGGCC | 62 | TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY |
| QY | 768 | TCAACTCATATCGTGCACCTGTCGCGGACCCCGGGGCGGTGTCCTCCGTCGAGAAC | 827 | TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER |
| Db | 63 | TGACCTTAAGTATCTCCACTTGGTGGCTGATCCCGGGCGGTGGCAGTTTCAGCATCC | 122 | FILE REFERENCE: 210121.512 |
| QY | 828 | GCACAAAGGAGATCTCATGATTTGACAGTCCGATTTGTATGGGGCAGCATGACGAG | 883 | CURRENT APPLICATION NUMBER: US/09/833.790 |
| Db | 123 | GCTCGCGCCACGCGCTCATCCGTAGAGACCTTCACAGTGGTGGCGACGCGAGACCG | 178 | CURRENT FILING DATE: 2001-04-11 |
| <p>RESULT 14</p> <p>US-09-833-790-207/c</p> <p>Sequence 207, Application US/09833790</p> <p>Patent No. US20020068288A1</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Lodes; Michael J.</p> <p>APPLICANT: Wang, Tongtong</p> <p>APPLICANT: Secrist, Heather</p> <p>APPLICANT: Momach, Radooh</p> <p>APPLICANT: Indrias, Carol Y.</p> <p>APPLICANT: Fan, Liqun</p> <p>TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY</p> <p>TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER</p> <p>FILE REFERENCE: 210121.512</p> <p>CURRENT APPLICATION NUMBER: US/09/833.790</p> <p>CURRENT FILING DATE: 2001-04-11</p> <p>NUMBER OF SEQ ID NOS: 440</p> <p>SOFTWARE: FastSeq for Windows Version 4.0</p> <p>SEQ ID NO 207</p> <p>LENGTH: 395</p> <p>TYPE: DNA</p> <p>ORGANISM: Homo sapien</p> <p>US-09-833-790-207</p> | | | | |
| QY | 205 | GCTCTGCTGTTTCTTGCTTTCCCAATGGCCATCTTGCTCTATTTCTCCACATGTACG | 264 | Query Match |
| Db | 285 | GGCCAGGCGCTTGTATGTCCGAGTGTGCATCTTGGGTTTCTGGGCCATTTGGC | 226 | Best Local Similarity 49.5%; Pred. No. 0.013; |
| QY | 265 | CCACACATCAGTCCCTCTCTATGAAGACACAGCCGAGAGCATGCTGCTTCT | 324 | Matches 110; Conservative 0; Mismatches 112; Indels 0; Gaps 0; |
| Db | 225 | CCGCTGCCCGGGACACACATGAAAGGATTCATGGGCCGCTTGACGGGATCCGGCC | 166 | |
| QY | 325 | GTCCTCTGAGCGCTTGGCTTTCTTTTGTGGGCGAGCTTTTGGGACGACCCATATG | 384 | |
| Db | 165 | GTTTTCTGTATGCCCGCGCCGCCCGCGGTGAGATGCCCGCCCGCCCGAAGT | 106 | |
| QY | 385 | TTTCTACGTGATGAGACCGGCTGGGACGATGTGATGACCTT | 426 | |

Db 1711 TCACCTTCACACGATCAACAGAGGTGGAGAGTTTGTCTACAGCCATGCCCGCTGG 1770
QY 1254 GCTACCGCAGCTCA 1268
Db 1771 GCTATGACCGGCTCA 1785

RESULT 9
US-09-735-705-63

Sequence 63, Application US/09735705
Patent No. US20020052329A1
GENERAL INFORMATION:

APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
APPLICANT: Fanger, Neil

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.455C14
CURRENT APPLICATION NUMBER: US/09/735.705

CURRENT FILING DATE: 2000-12-12
NUMBER OF SEQ ID NOS: 419

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63

LENGTH: 731

TYPE: DNA
ORGANISM: Homo sapien

FEATURE:
NAME/KEY: misc.feature

LOCATION: (1)...(731)
OTHER INFORMATION: n = A,T,C or G

US-09-735-705-63

Query Match 3.1% Score 62.6; DB 10; Length 731;
Best Local Similarity 58.2%; Pred. No. 2.8e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 702 TGGTGTCAAGAGAGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTAAGACC 761
Db 3 TAGTCATAAAGGGTGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTAAGACC 761
QY 762 CCTCCCTCAACCTGCATATGATGACCTGGTGGGAGCCCGCGCGCTGTCGTTCCC 821
Db 63 CGGCGCTGGAGCTCAAGCTCATCTGCTGTGCTGTATCCCCCGCGGTGGCGAGTTCC 122
QY 822 GAGAAGCACAAGAGGAGTCTCATGTATGACAGTCCGATTTGATGGGCGACATGAGC 881
Db 123 GGATCCCGCTGGCGCAGGCTCATCCGTGAGAGCTACAGGTGTGGCGAGCCGAGACC 182
QY 882 AGAAGCTCA 890
Db 183 GCGAGCTCA 191

RESULT 10
US-09-850-716A-63
Sequence 63, Application US/09850716A
Patent No. US20020115139A1
GENERAL INFORMATION:
APPLICANT: Kalos, Michael D.
APPLICANT: McNeill, Patricia D.
APPLICANT: Retter, Marc W.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.455C15
CURRENT APPLICATION NUMBER: US/09/850.716A
CURRENT FILING DATE: 2001-05-07
NUMBER OF SEQ ID NOS: 440
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(731)
OTHER INFORMATION: n = A,T,C or G

US-09-850-716A-63

Query Match 3.1% Score 62.6; DB 10; Length 731;
Best Local Similarity 58.2%; Pred. No. 2.8e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 702 TGGTGTCAAGAGAGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTAAGACC 761
Db 3 TAGTCATAAAGGGTGTGGCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTAAGACC 761
QY 762 CCTCCCTCAACCTGCATATGATGACCTGGTGGGAGCCCGCGCGCTGTCGTTCCC 821
Db 63 CGGCGCTGGAGCTCAAGCTCATCTGCTGTGCTGTATCCCCCGCGGTGGCGAGTTCC 122
QY 822 GAGAAGCACAAGAGGAGTCTCATGTATGACAGTCCGATTTGATGGGCGACATGAGC 881
Db 123 GGATCCCGCTGGCGCAGGCTCATCCGTGAGAGCTACAGGTGTGGCGAGCCGAGACC 182
QY 882 AGAAGCTCA 890
Db 183 GCGAGCTCA 191

RESULT 11

US-09-897-778-63

Sequence 63, Application US/09897778
Patent No. US20020147143A1
GENERAL INFORMATION:

APPLICANT: Wang, Tongtong
APPLICANT: Mamerakis, Margarita
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darriick
APPLICANT: Watanabe, Yoshihiro
APPLICANT: Henderson, Robert A.
APPLICANT: Peckham, David W.
APPLICANT: Fanger, Neil

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.455C16
CURRENT APPLICATION NUMBER: US/09/897.778

CURRENT FILING DATE: 2001-06-28
NUMBER OF SEQ ID NOS: 467

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 63

LENGTH: 731

TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

LOCATION: 237, 248, 263, 288, 312, 317, 323, 326, 337, 352, 362, 370,
LOCATION: 377, 400, 411, 414, 434, 436, 446, 457, 473, 486, 497, 498,
LOCATION: 502, 512, 531, 546, 554, 565, 588, 597, 608, 611,
LOCATION: 613, 615, 627, 632, 640, 641, 644, 654, 660, 663, 665

OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc.feature

LOCATION: 671, 678, 692, 697, 698, 699, 704, 705, 712, 714, 717, 718,
LOCATION: 719, 722, 725, 730, 731

OTHER INFORMATION: n = A,T,C or G

US-09-897-778-63

Db 48368 CGGAGCCGCTGCGAGAAATCCGCTCTACGCCCTTCACTGGGCTCAGTCTCAGGCCA 48427
QY 1082 CATCTTCAG 10950
Db 48428 CAGCTCGAG 48436

RESULT 7
US-09-919-580-264
; Sequence 264, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; CURRENT FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264
; LENGTH: 389
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-264

Query Match 14.1%; Score 287.4; DB 10; Length 389;
Best Local Similarity 99.7%; Pred. No. 1.2e-74;
Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1729 GAGCTACCAACGATTTTCCACAGAGATGCAAAATTTGAGCCCTTGAGTTCCACATGAGAT 1788
Db 10 GACCTCCACGACATTTTCCACAGAGATGCAAAATTTGAGCCCTTGAGTTCCACATGAGAT 69
QY 1789 TCAGGAGAGAGAGTGGGAGACAGATGAGTGGCTTATGAGTATGAGTATGAGTATGAGTAT 1848
Db 70 TCAGGAGAGAGAGTGGGAGACAGATGAGTGGCTTATGAGTATGAGTATGAGTATGAGTAT 129
QY 1849 TCGGTAAATCAGAAATATGAAACAAATCTCTGCACAAAGAGAGAGTCTTAAGTTTACA 1908
Db 130 TCGGTAAATCAGAAATATGAAACAAATCTCTGCACAAAGAGAGAGTCTTAAGTTTACA 189
QY 1909 GGGTGGCCGGGCTGATTTGATATATCACTGCGCTTGCATTTTCCCATACATAGAGA 1968
Db 190 GGGTGGCCGGGCTGATTTGATATATCACTGCGCTTGCATTTTCCCATACATAGAGA 249
QY 1969 CTTTGACCTGTGAAGTGGCATCTGTTAATTAATAAATTCACCAATAG 2017
Db 250 CTTTGACCTGTGAAGTGGCATCTGTTAATTAATAAATTCACCAATAG 298

RESULT 8
US-09-833-790-244
; Sequence 244, Application US/09833790
; Patent No. US2002006828A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Monamath, Radonh
; APPLICANT: Indrias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244
; LENGTH: 2393

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-790-244

Query Match 5.9%; Score 119.4; DB 10; Length 2393;
Best Local Similarity 49.1%; Pred. No. 1.2e-24;
Matches 508; Conservative 0; Mismatches 446; Indels 81; Gaps 4;

QY 315 TGCTGTTCTGTCTCTCTCTGCGCCCTGAGCTCTCTTTTGCGGCGAGCTTTTGCGCAGC 374
Db 751 TGTAGTGTTCACACGAGCGGCTCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 810
QY 375 ACCCAGATGTTTCTACCTGATGAGAGCCCGCTGCGAGCTGTGATGACCTCAAGCAGA 434
Db 811 ATCCGAGGTGTTCTTTCTCTACAGAGCAGTGTGCAATGATGCAAAACTGTATCCG 870
QY 435 GCACGCGCTGATGCTGACATGCGTGTGCGGAGTGTATACGCGCTCTCTCTCTCTCTCT 494
Db 871 GCGAGCGCTTCCCTGACGAGGCGGAGCGGAGACATGATGAGCGCTCTTACCGCTCG 930
QY 495 ACATGAGCTCTTTGATGCTTACATGAACTGTTCCTCCGAGAGC-----AGT 542
Db 931 ACCTCTCTGTCTTCCAGTGTATAGCCCGCGGCGAGCGGCGGCGCAACTCACCACGC 990
QY 543 CCAGCCTCTTCTAGTGGAGAACAGCGGCGCTGTGTCTTGTGCACTCTGTGACATCA 602
Db 991 TGGGCACTTTCGCGCGAGCGACCAACAGTGTGTCTGTCTGTCTGTCTGTCTGTCTGT 1050
QY 603 TCCACAGATGAATCATCCCGCGGCTCACTGAGGCTC---CTGTGACATCAACGC 659
Db 1051 ACCGAGAGAGTGTGCGGCTGTGTGTGAGACGCGGTGTGCAAGAGATGCGCGCCACGC 1110
QY 660 CTTTATGAGTGTGAGAGAGCGCTGCGCTCTCTTACAGCCAGCTGTGCTGAGAGTGTG 719
Db 1111 GCTTGGCGGCTTTCGAGAGAGTGTGCGCAAGTATCCGCACTATGATTAAGGTGTG 1170
QY 720 GCTTCTTCACTGTGAGTCTCTTACCCGCTGCTGAAAGACCCCTCTCTCAACTGTCATA 779
Db 1171 GCTTCTTCACTGTGAGTCTCTTACCCGCTGCTGAAAGACCCGCTCTCTCAACTGTCATA 1230
QY 780 TCGTGTACTGTGTGCGGAGACCCCGCGCTGTCTGCTTCCGAGAGAGCACAAGGAGG 839
Db 1231 TCAATCACTGTGTGCTGTATCCCGCGGCTGTGCGAGTGTGATGATCGCTGCGCAGC 1290
QY 840 ATTCATGATTTGACATCTCCATTTGTGATG----- 868
Db 1291 GCTTCACTGTGAGAGCTTCAAGTGTGTGCGAGCGGAGCCGAGCTCACCGCATGC 1350
QY 869 -----GGGACAGATGAGCAGAACTCAAGAGAGG----- 899
Db 1351 CTTTCTTGAAGCGCGCGGCGCACAGCTTGGCGCAGAGAGGCGCTGGCGCGCGCCG 1410
QY 900 -----ACCAACCTACTATGTATGATGAGTCAATCTGCAAAAGCCAGTGAATCTACA 953
Db 1411 CAGACTACACGCTCTGCGGCGTATGAGGTATCTGCAATGATATGCTTGAAGCGCTGC 1470
QY 954 AGACCATCAGTCTTGTGCCAAGGCCCTGAGAGAGCTTCTGTGCGCTATAGAG 1013
Db 1471 AGACAGCCCTGAGCCCGCTGACTGTGAGGCGCACATCGGTGTGAGGTAGAG 1530
QY 1014 ACCTGAGTGTGAGCCCGCTGAGCCAGCTTCCGAGTATGATGATGATGATGATGAT 1073
Db 1531 ACCTGAGTGTGAGCCCGCTGAGCCAGCTTCCGAGTATGATGATGATGATGATGAT 1590
QY 1074 TCTTGGCCATCTTTCAGACCTGTGATGATGATGATGATGATGATGATGATGATGAT 1133
Db 1591 TGAAGCCCGAATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1650
QY 1134 AGCTTTCACACAAATGCGAGGATGCTTATGATGATGATGATGATGATGATGATGAT 1193
Db 1651 AGCTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1253
QY 1194 TGCCCTATGAAAGGTTTCTGACTTCAGAAAGCTGTGCGATGCCATGATTTGTCTG 1253


```

: APPLICANT: Sectist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE REFERENCE: 210121.552
: CURRENT APPLICATION NUMBER: US/09/919,580
: CURRENT FILING DATE: 2001-07-30
: NUMBER OF SEQ ID NOS: 934
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 194
: LENGTH: 517
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-919-580-194

Query Match      24.9%; Score 506.4; DB 10; Length 517;
Best Local Similarity 99.8%; Pred. No. 4.6e-139;
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1352 GAAGGCTTCTGCGACCTGCTGTCAGCCCTGACATCTTCTCTGATGCTTGTGAGCCT 1411
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 10 GAGGCTTCTGCTGCGACCTGCTGTCAGCCCTGACATCTTCTCTGATGCTTGTGAGCCT 69

QY 1412 TGCCATCATCTGAGGCTTAACTACTATGCTGTGTGGTATCACATGAGTGTGAGTTG 1471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 70 TGCCATCATCTGAGGCTTAACTACTATGCTGTGTGGTATCACATGAGTGTGAGTTG 129

QY 1472 TCCACAGCTGCTGACAGGAGGAGGACTTTGTGTCATGCTTGTGTCAGAAACAGACTG 1531
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 130 TCCACAGCTGCTGACAGGAGGAGGACTTTGTGTCATGCTTGTGTCAGAAACAGACTG 189

QY 1532 GGGACCTTATGTGAGGAGGACATCCACAGTGAAGAGGATGCTTCTTCTTTT 1591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 190 GGGACCTTATGTGAGGAGGACATCCACAGTGAAGAGGATGCTTCTTCTTTT 249

QY 1592 CTTATCTTCTGCTGCTGGGAGACTTGTAGAGACTTGTGGCCTGGAGGCTTATTAAGC 1651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 250 CTTATCTTCTGCTGCTGGGAGACTTGTAGAGACTTGTGGCCTGGAGGCTTATTAAGC 309

QY 1652 GACACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1711
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 310 GACACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 369

QY 1712 ATGATCTTTCACCAAGAGGTCACAGCAATTTTCCACAGAGATGCAATTTGAGGCC 1771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 370 ATGATCTTTCACCAAGAGGTCACAGCAATTTTCCACAGAGATGCAATTTGAGGCC 429

QY 1772 TGGAGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGTTGATGCTTATGATGAG 1831
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 430 TGGAGTTCCAGTGTGATTCAGAGAGAGTGGAGACAGTTGATGCTTATGATGAG 489

QY 1832 CTGACCATCACAGCTATCGGTAATCAG 1859
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 490 CTGACCATCACAGCTATCGGTAATCAG 517

RESULT 4
US-09-998-598-2595
: Sequence 2595, Application US/09998598
: Patent No. US20020150922A1
: GENERAL INFORMATION:
: APPLICANT: Stolk, John A.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Chenault, Ruth A.
: APPLICANT: Mesgher, Madeleine Joy
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: FILE REFERENCE: 210121.561
: CURRENT APPLICATION NUMBER: US/09/998,598
: CURRENT FILING DATE: 2001-11-16
: NUMBER OF SEQ ID NOS: 2606
: SOFTWARE: Corixa Invention Disclosure Database
: SEQ ID NO 2595
: LENGTH: 505
```

```

: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-998-598-2595

Query Match      24.9%; Score 505; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.2e-138;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1355 GCGTTGCTGCGACCTGCTGTCAGCCCTGACATCTTCTCTGATGCTTGTGAGCCTG 1414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 GCGTTGCTGCGACCTGCTGTCAGCCCTGACATCTTCTCTGATGCTTGTGAGCCTG 60

QY 1415 CTACATCTCTGAGCCTTAACTACTATGCTGTGTGGTATCACATGAGTGTGATGCTC 1474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 CTACATCTCTGAGCCTTAACTACTATGCTGTGTGGTATCACATGAGTGTGATGCTC 120

QY 1475 AACGTGCTCAAGCAGAGAGACTTTTGTGTGATGCTTGTGTGATGCTTGTGATGCTG 1534
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AACGTGCTCAAGCAGAGAGACTTTTGTGTGATGCTTGTGTGATGCTTGTGATGCTG 180

QY 1535 AACCTATGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1594
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AACCTATGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240

QY 1595 GATCTTCTGCTGCTGGGAGACTTTCAGAGACTTGTGGCCTGGAGGCTTATTAAGCAG 1654
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 GATCTTCTGCTGCTGGGAGACTTTCAGAGACTTGTGGCCTGGAGGCTTATTAAGCAG 300

QY 1655 ACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1714
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 ACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360

QY 1715 GATCTTTCACCAAGAGGTCACAGCAATTTTCCACAGAGATGCAATTTGAGGCC 1774
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GATCTTTCACCAAGAGGTCACAGCAATTTTCCACAGAGATGCAATTTGAGGCC 420

QY 1775 AGTTCACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1834
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 AGTTCACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480

QY 1835 GACCATCACAGCTATCGGTAATCAG 1859
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 GACCATCACAGCTATCGGTAATCAG 505

RESULT 5
US-09-927-602-1
: Sequence 1, Application US/09927602
: Patent No. US20020061562A1
: GENERAL INFORMATION:
: APPLICANT: Akama, Michiko N.
: APPLICANT: Tomoya O.
: TITLE OF INVENTION: Methods of Treating Macular Corneal
: FILE REFERENCE: P-LJ 4852
: CURRENT APPLICATION NUMBER: US/09/927,602
: CURRENT FILING DATE: 2001-08-09
: PRIOR APPLICATION NUMBER: US 09/638,211
: PRIOR FILING DATE: 2000-08-11
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 2544
: TYPE: DNA
: ORGANISM: Homo Sapien
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (693)...(1877)
: US-09-927-602-1

Query Match      20.4%; Score 414.4; DB 10; Length 2544;
Best Local Similarity 64.5%; Pred. No. 1.6e-111;
Matches 654; Conservative 0; Mismatches 351; Indels 9; Gaps 2;
```

Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGTGGAGGCGAGATGGCTCAGTCTGGGGGAAATGCTTCATTTGGTCTCCAG 60
Db 1 GCGTGGAGGCGAGATGGCTCAGTCTGGGGGAAATGCTTCATTTGGTCTCCAG 60

QY 61 CCCAGCTCAAGAGCTCTCCCAACCCCTTGAATCAGCAGTTAAAGCTTTACTTTCA 120
Db 61 CCCAGCTCAAGAGCTCTCCCAACCCCTTGAATCAGCAGTTAAAGCTTTACTTTCA 120

QY 121 CAGCTTCCTGGGAGGAGTGGTCTTCTCAAGCCCTTGGCAAGTCTTCCACTCAGAC 180
Db 121 CAGCTTCCTGGGAGGAGTGGTCTTCTCAAGCCCTTGGCAAGTCTTCCACTCAGAC 180

QY 181 AATGCTACTGGCTTAAATAAATGAAGCTCCTGCTTTCTGATTTCCAGATGGCATCTT 240
Db 181 AATGCTACTGGCTTAAATAAATGAAGCTCCTGCTTTCTGATTTCCAGATGGCATCTT 240

QY 241 GGCCTATATTTCCACATGTCAGCAGCCACATACATAGCTCCTGCTCTATGAAGCAGCC 300
Db 241 GGCCTATATTTCCACATGTCAGCAGCCACATACATAGCTCCTGCTCTATGAAGCAGCC 300

QY 301 CGAGGCGATGCGAGCTGCTGTCTCTCTGCGGCTCTGCTCTTCTTGTGGGAGCA 360
Db 301 CGAGGCGATGCGAGCTGCTGTCTCTCTGCGGCTCTGCTCTTCTTGTGGGAGCA 360

QY 361 GCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGAGCCGCTGGCAGCTGTGGAT 420
Db 361 GCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGAGCCGCTGGCAGCTGTGGAT 420

QY 421 GACCTTCAGAGAGCAGCCGCTGATGCTGACATGCTGCTGAGGATGATGAGGAGG 480
Db 421 GACCTTCAGAGAGCAGCCGCTGATGCTGACATGCTGCTGAGGATGATGAGGAGG 480

QY 481 CGTCTCTTGTGGCAGATGAGCGTCTTGTGATGCTACATGGAACCTGTGCCCCGAGACA 540
Db 481 CGTCTCTTGTGGCAGATGAGCGTCTTGTGATGCTACATGGAACCTGTGCCCCGAGACA 540

QY 541 GTCCAGCCTCTTTCAGTGGGAGAGACAGCCGCGCTGCTGCTGACCTGCTGACAT 600
Db 541 GTCCAGCCTCTTTCAGTGGGAGAGACAGCCGCGCTGCTGCTGACCTGCTGACAT 600

QY 601 CATCCCAAGAGATGAATCATCCCGCGGCTCACTGAGGCTCCTGCGAGTCAACAGCC 660
Db 601 CATCCCAAGAGATGAATCATCCCGCGGCTCACTGAGGCTCCTGCGAGTCAACAGCC 660

QY 661 CTTTGGAGTGGTGGAGAGGCTGCGGCTGCTTACAGCAGCTGCTCAAGAGAGGTCG 720
Db 661 CTTTGGAGTGGTGGAGAGGCTGCGGCTGCTTACAGCAGCTGCTCAAGAGAGGTCG 720

QY 721 CTCTTCAACCTGAGTCCCTTACCCGCTGCTGAAGAGACCCCTCCCTCACTGATAT 780
Db 721 CTCTTCAACCTGAGTCCCTTACCCGCTGCTGAAGAGACCCCTCCCTCACTGATAT 780

QY 781 CGTGCACCTGCTCGGGAGCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 781 CGTGCACCTGCTCGGGAGCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840

QY 841 TCTCATGATGAGAGTGCATGATGATGGGAGCAGATGAGCAGAACTCAAGAGAGAGA 900
Db 841 TCTCATGATGAGAGTGCATGATGATGGGAGCAGATGAGCAGAACTCAAGAGAGAGA 900

QY 901 CCAACCTTACTATGATGAGTGCATGCTGCAAGAGCAGCTGAGATCTACAGAGCAT 960
Db 901 CCAACCTTACTATGATGAGTGCATGCTGCAAGAGCAGCTGAGATCTACAGAGCAT 960

QY 961 CCAAGCTTCTGCAAGGCGCTGCAAGAGCAGTCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db 961 CCAAGCTTCTGCAAGGCGCTGCAAGAGCAGTCTGCTGCTGCTGCTGCTGCTGCTG 1020

QY 1021 TCGAGCCCTGTGCGCCAGACTTCCCGAATGATGAATTCGTGGATTCGATTCCTGCC 1080
Db 1021 TCGAGCCCTGTGCGCCAGACTTCCCGAATGATGAATTCGTGGATTCGATTCCTGCC 1080

QY 1081 CCATCTTCAGACCTGGGTGCTATACATCACCCGAGGCAAGGGCATGGGTGACCAAGCTTT 1140
Db 1081 CCATCTTCAGACCTGGGTGCTATACATCACCCGAGGCAAGGGCATGGGTGACCAAGCTTT 1140

QY 1141 CCACACAAATGCCAGGAGTCCCTTAATGCTCCAGAGGCTGCGGCTGGCTTTGGCCCTA 1200
Db 1141 CCACACAAATGCCAGGAGTCCCTTAATGCTCCAGAGGCTGCGGCTGGCTTTGGCCCTA 1200

QY 1201 TGAAGGCTTTCAGACTTCAGAAAGCTGTGGCATGCTCCATGAATTTGCTGGGCTACCG 1260
Db 1201 TGAAGGCTTTCAGACTTCAGAAAGCTGTGGCATGCTCCATGAATTTGCTGGGCTACCG 1260

QY 1261 CCAGCTGAGTCTGACCAAGAGAGAACTGCTGCTGAGTCTGCTGCTGCTGCTGCTG 1320
Db 1261 CCAGCTGAGTCTGACCAAGAGAGAACTGCTGCTGAGTCTGCTGCTGCTGCTGCTG 1320

QY 1321 TGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAGGCTTGTGCTCCACCTGTGTGAGCC 1380
Db 1321 TGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAGGCTTGTGCTCCACCTGTGTGAGCC 1380

QY 1381 TCACTGACTTCTCTGAATGCTCTGAGCCTTGCCTACATCTCTGAGCTTAACTACATG 1440
Db 1381 TCACTGACTTCTCTGAATGCTCTGAGCCTTGCCTACATCTCTGAGCCTTAACTACATG 1440

QY 1441 TCTGTGGATTCACACTGAGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Db 1441 TCTGTGGATTCACACTGAGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500

QY 1501 GTGTCCATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
Db 1501 GTGTCCATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560

QY 1561 CAGTGAAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
Db 1561 CAGTGAAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620

QY 1621 AGACTTGTGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
Db 1621 AGACTTGTGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680

QY 1681 ACCCTCCCTGCAACCTTGGCCCAATGGGAGTGGATCTTTCACCAAGAGTCCACAGC 1740
Db 1681 ACCCTCCCTGCAACCTTGGCCCAATGGGAGTGGATCTTTCACCAAGAGTCCACAGC 1740

QY 1741 ATTTTCACAGAGATGCAAAATCTGAGCCCTTGGAGTCTCCAGTGGATTCAGAGAGAA 1800
Db 1741 ATTTTCACAGAGATGCAAAATCTGAGCCCTTGGAGTCTCCAGTGGATTCAGAGAGAA 1800

QY 1801 GTGGGAGCAAGGTTGAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
Db 1801 GTGGGAGCAAGGTTGAGTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860

QY 1861 AATATGAACAAATCTGCAACAAAGAGCAGCTCTTAAGTTTCAAGAGTGGCTGGGC 1920
Db 1861 AATATGAACAAATCTGCAACAAAGAGCAGCTCTTAAGTTTCAAGAGTGGCTGGGC 1920

QY 1921 TGCATTTGAATATCACTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Db 1921 TGCATTTGAATATCACTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980

QY 1981 AAGCTGCATCTGTTAATACTAAATTTCCCAATAGAGAGAGAGAGAGAGAGAGAG 2032
Db 1981 AAGCTGCATCTGTTAATACTAAATTTCCCAATAGAGAGAGAGAGAGAGAGAGAG 2032

RESULT 3
US-09-919-580-194
; Sequence 194, Application us/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyile, Ruth
; APPLICANT: Xu, Jiangchun

QY 241 GGCCTATTTCTCCAGTATGACAGCCACACATCAGCTCCCTGCTATATGAAGGACACACC 300
 |||||
 Db 241 GGCCTATTTCTCCAGTATGACAGCCACACATCAGCTCCCTGCTATATGAAGGACACACC 300
 QY 301 CGAGCCCATGACAGCTGTGTGTCTGTCTCTCTGCGCTCTGTGCTCTTTTGTGGGACA 360
 |||||
 Db 301 CGAGCCCATGACAGCTGTGTGTCTGTCTCTCTGCGCTCTGTGCTCTTTTGTGGGACA 360
 QY 361 GCTTTTGGGACGACCCAGATGTTTCTACCTGATGAGACCCGCTGGACAGTGTGAT 420
 |||||
 Db 361 GCTTTTGGGACGACCCAGATGTTTCTACCTGATGAGACCCGCTGGACAGTGTGAT 420
 QY 421 GACCTTCAAGAGAGACCCGCTGATGCTGACATGCTGTGGGGATCTGATATCGGGC 480
 |||||
 Db 421 GACCTTCAAGAGAGACCCGCTGATGCTGACATGCTGTGGGGATCTGATATCGGGC 480
 QY 481 CGCTTCTGTGCGACATGAGGCTTTTGATGCTGACATGAGAACTGGTCCCGGAGACA 540
 |||||
 Db 481 CGCTTCTGTGCGACATGAGGCTTTTGATGCTGACATGAGAACTGGTCCCGGAGACA 540
 QY 541 GTCCAGCCTCTTCTAGTGGGAGAGACCGGGCCTGTGTCTGTGACCTGCTGTGACAT 600
 |||||
 Db 541 GTCCAGCCTCTTCTAGTGGGAGAGACCGGGCCTGTGTCTGTGACCTGCTGTGACAT 600
 QY 601 CATCCCAAGATGAAATCATCCCCGGGCTACTGACAGGCTCTGTGACATCAACAGCC 660
 |||||
 Db 601 CATCCCAAGATGAAATCATCCCCGGGCTACTGACAGGCTCTGTGACATCAACAGCC 660
 QY 661 CTTGAGGGTGGAGAAAGGCTGCGGCTCTCAAGCCAGTGTGTGATGAAGAGAGTGG 720
 |||||
 Db 661 CTTGAGGGTGGAGAAAGGCTGCGGCTCTCAAGCCAGTGTGTGATGAAGAGAGTGG 720
 QY 721 CTCTTCAACCTGACAGTCTCTCAACCCGCTGTGAAGAACCCCTCTCACTGATAT 780
 |||||
 Db 721 CTCTTCAACCTGACAGTCTCTCAACCCGCTGTGAAGAACCCCTCTCACTGATAT 780
 QY 781 CGTGACCTGTGACAGGACCCCGGGCCTGTGTGCTTCCGAGAGACGACAAAGGAGA 840
 |||||
 Db 781 CGTGACCTGTGACAGGACCCCGGGCCTGTGTGCTTCCGAGAGACGACAAAGGAGA 840
 QY 841 TCTCATGATGACAGT 900
 |||||
 Db 841 TCTCATGATGACAGT 900
 QY 901 CCAACCTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 |||||
 Db 901 CCAACCTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 QY 961 CCAAGCTTGTGCGAGGCTGTGAGAGACGCTACCTGTGTGTGTGTGTGTGTGTGTGT 1020
 |||||
 Db 961 CCAAGCTTGTGCGAGGCTGTGAGAGACGCTACCTGTGTGTGTGTGTGTGTGTGTGT 1020
 QY 1021 TCAGAGCCCTGTGCGAGGCTGTGAGAGACGCTACCTGTGTGTGTGTGTGTGTGTGT 1080
 |||||
 Db 1021 TCAGAGCCCTGTGCGAGGCTGTGAGAGACGCTACCTGTGTGTGTGTGTGTGTGTGT 1080
 QY 1081 CCAATCTTCAAGCTGT 1140
 |||||
 Db 1081 CCAATCTTCAAGCTGT 1140
 QY 1141 CCAACACAAATGCGAGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1200
 |||||
 Db 1141 CCAACACAAATGCGAGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1200
 QY 1201 TGAAGAGCTTCTGAGCTTGAAGAGCTGTGCGATGCCATGAATTTGTGGGTACCG 1260
 |||||
 Db 1201 TGAAGAGCTTCTGAGCTTGAAGAGCTGTGCGATGCCATGAATTTGTGGGTACCG 1260
 QY 1261 CCAAGTCAATCTGAGACAGACAGAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320
 |||||
 Db 1261 CCAAGTCAATCTGAGACAGACAGAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320

QY 1321 TGTCCCTGAGCAATTCACATTAAGAGGCTTGAAGAAGCTTGTGCTGCCACTGTGTGAGCC 1380
 |||||
 Db 1321 TGTCCCTGAGCAATTCACATTAAGAGGCTTGAAGAAGCTTGTGCTGCCACTGTGTGAGCC 1380
 QY 1381 TCAGTCACTTCTGTGATGATGCTTGTGAGCCCTGTGACATCTGTGAGCCCTTAATACATG 1440
 |||||
 Db 1381 TCAGTCACTTCTGTGATGATGCTTGTGAGCCCTGTGACATCTGTGAGCCCTTAATACATG 1440
 QY 1441 TCTGTGGGATACACATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
 |||||
 Db 1441 TCTGTGGGATACACATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
 QY 1501 TCTGTGGGATACACATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
 |||||
 Db 1501 TCTGTGGGATACACATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
 QY 1561 CAGTGAACAGAGGATGT 1620
 |||||
 Db 1561 CAGTGAACAGAGGATGT 1620
 QY 1621 AGACTTGTGCGCTGAGAGGCTTATTAAGACAGACAGATGATGATGATGATGATGATGAT 1680
 |||||
 Db 1621 AGACTTGTGCGCTGAGAGGCTTATTAAGACAGACAGATGATGATGATGATGATGATGAT 1680
 QY 1681 ACCTTGTGCGCTGAGAGGCTTATTAAGACAGACAGATGATGATGATGATGATGATGAT 1740
 |||||
 Db 1681 ACCTTGTGCGCTGAGAGGCTTATTAAGACAGACAGATGATGATGATGATGATGATGAT 1740
 QY 1741 ATTTTCCAGAGATGCAAAATTTGAGGCTTGAAGATTTCCAGATTTCAAGAGAGAA 1800
 |||||
 Db 1741 ATTTTCCAGAGATGCAAAATTTGAGGCTTGAAGATTTCCAGATTTCAAGAGAGAA 1800
 QY 1801 GTGGAAACAGGTTGAGAGGCTTATTAAGACAGACAGATGATGATGATGATGATGATGAT 1860
 |||||
 Db 1801 GTGGAAACAGGTTGAGAGGCTTATTAAGACAGACAGATGATGATGATGATGATGATGAT 1860
 QY 1861 AATATGAACAAATCTGTGACAAAGAGAGAGCTTTAAGTTACAGAGGCTGCTGGGC 1920
 |||||
 Db 1861 AATATGAACAAATCTGTGACAAAGAGAGAGCTTTAAGTTACAGAGGCTGCTGGGC 1920
 QY 1921 TGCATTTAATATCACTTCCCTGCTGCAATTTCCATGACATTAAGAGCTTGAAGCTGTG 1980
 |||||
 Db 1921 TGCATTTAATATCACTTCCCTGCTGCAATTTCCATGACATTAAGAGCTTGAAGCTGTG 1980
 QY 1981 AAGCTGCATCTGTATTAATCTAAATTTCCAAATTAAGAAAAA 2032
 |||||
 Db 1981 AAGCTGCATCTGTATTAATCTAAATTTCCAAATTAAGAAAAA 2032

RESULT 2

US-09-816-825-1
 ; Sequence 1, Application US/09816825
 ; Patent No. US20010051370A1
 GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven B.
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CON
 ; CURRENT APPLICATION NUMBER: US/09/816, 825
 ; PRIORITY FILING DATE: 2001-03-22
 ; PRIORITY APPLICATION NUMBER: 09/045, 284
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO: 1
 ; LENGTH: 2032
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-816-825-1
 Query Match 100.0%; Score 2032; DB 10; Length 2032;
 Best Local Similarity 100.0%; Pred. No. 0;

| | Query Match | Similarity | 100.0% | Score | 2032 | DB | 9 | Length | 2032 |
|----|-------------|---|--------------|-------|------------|----|--------|--------|------|
| | Best Local | Similarity | 100.0% | Pred. | NC.0 | | | | |
| | Matches | 2032 | Conservative | 0 | Mismatches | 0 | Indels | 0 | Gaps |
| QY | 1 | GGCTCGAGGCGAGATGCCCTCAGTCTGTGGGGGAAAAATGCTTCATATTGGTCTTCCAG | 60 | | | | | | |
| Db | 1 | GGCTCGAGGCGAGATGCCCTCAGTCTGTGGGGGAAAAATGCTTCATATTGGTCTTCCAG | 60 | | | | | | |
| QY | 61 | CCCACTCAAGAGATCTGCCCAACCCCTTGAGTCTCAGACAGTGTAAAGCTGTACTTTCA | 120 | | | | | | |
| Db | 61 | CCCACTCAAGAGATCTGCCCAACCCCTTGAGTCTCAGACAGTGTAAAGCTGTACTTTCA | 120 | | | | | | |
| QY | 121 | CAGCTTCGAGGAGGAGTCTTCTTCAACCCGCTCTTCACAAGTCTTCCACTTCAGAC | 180 | | | | | | |
| Db | 121 | CAGCTTCGAGGAGGAGTCTTCTTCAACCCGCTCTTCACAAGTCTTCCACTTCAGAC | 180 | | | | | | |
| QY | 181 | AATGTACTTGCTAAAAAATGAAGCTCTCTGCTGTTCGTGTTTCCACATGCGCATCTT | 240 | | | | | | |
| Db | 181 | AATGTACTTGCTAAAAAATGAAGCTCTCTGCTGTTCGTGTTTCCACATGCGCATCTT | 240 | | | | | | |

This Page Blank (uspto)


```
LOCATION: (411)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (414)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (434)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (436)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (445)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (457)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (473)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (486)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (497)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (498)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (502)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (512)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (531)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (546)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (554)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (563)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (565)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (588)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (597)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (606)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (611)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (613)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (615)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (627)
```

```
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (632)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (640)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (641)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (644)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (654)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (660)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (663)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (665)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (671)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (678)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (692)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (697)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (698)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (699)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (704)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (705)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (712)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (714)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (717)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (718)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (719)
OTHER INFORMATION: where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (723)
```

Query Match 3.1%; Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 1.9e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 702 TGGTGCATCAAGAGGTGGCGTTCTTCACCTGACGTCCCTACCGCGTGCATGAAGACC 761

Db 1174 GTGCCACCTCAAGCCCTGGGACGCTTGAGAGAGAGTGTGCGACAGTATCCGACAGGTGT 1233
QY 706 GCTCAGAGAGTGGCTTCTTCAACCTGCTCAGTCCCTCTACCCGCTGCTGAAGAAGCCCTC 765
Db 1234 TATCAAGGCGCGTGGCTTCTGAGTGTGTGTGGCGCGCTGTAAAGATCCAGC 1293
QY 766 CCTCACTGCTATCTGACAGCTGTCCGGGACCCCGGCGTGTCCGTTCCCGGAGA 825
Db 1294 CTTGAGCTCTCAAGTCACTCACTGCTGATCTGTCTGTGTGCGAGCTCCCGCAT 1353
QY 826 ACGCACAAGGAGATCTCATGATTGACAGTGCATTGTGATGGGCGGAGATGAGCAGA 885
Db 1354 CCGCTCGGCTACGCGCTCATCCGGAAGCCTACAGGTGTGGGAGAGCCGCGATCAAG 1413
QY 886 ACTCA 891
Db 1414 AGCCA 1419

RESULT 10
US-09-040-984-63
; Sequence 63, Application US/09040984
; Patent No. 6210883
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF LUNG CANCER
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,984
; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-282-6031
; TELEX:
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-040-984-63

Query March 3.1% Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 1.9e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 702 TGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTCTACCCGCTGCTGAAGAC 761
Db 3 TAGTCATAAAGGAGTGGCTTCTTCAAGAGTGGCGGCTTGGCGCCACCTGCTGAGAAC 62
QY 762 CTTCCCTCAACCTCATATGCTGACCTGGTCCGGGAGCCCGCGGCGGTGTTCCGTTCC 821
Db 63 CGGCCCTGAGACCTCAAGTCACTCATCTGTGTGTGTATCCCGCGCGGTGCGAGTTCC 122

QY 822 GAGACCGCACAAGAGATCTCATGATTGACAGTGCATGTGTGATGGGCGACATGAGC 881
Db 123 GATCCCTCGCCCGCCACGCTCATCCGTGAGAGCTTACAGGTGTGTGCGAGCCGAGACC 182
QY 882 AGAAGCTCA 890
Db 183 GCGAGCTCA 191

RESULT 11
US-09-123-912-63
; Sequence 63, Application US/09123912A
; Patent No. 6312695
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123,912A
; CURRENT FILING DATE: 1998-07-27
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 63
; LENGTH: 731
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (236)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (249)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (263)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (288)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (312)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (317)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (323)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (326)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (337)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (352)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (362)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (370)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (376)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (400)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base

Db 1771 GCTATGAGCGGTCA 1785

RESULT 8

US-09-263-023-1

; Sequence 1, Application US/09263023

; Patent No. 6037159

; GENERAL INFORMATION:

; APPLICANT: Uchimura, Kenji

; APPLICANT: Muramatsu, Hideki

; APPLICANT: Kadomatsu, Kenji

; APPLICANT: Kanagai, Reiji

; APPLICANT: Habuchi, Osami

; APPLICANT: Muramatsu, Takashi

; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND

; FILE REFERENCE: TOYAMA1.001AUS

; CURRENT APPLICATION NUMBER: US/09/263.023

; CURRENT FILING DATE: 1999-03-05

; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: JP 10-177844

; EARLIER FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 2150

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (470)...(1918)

US-09-263-023-1

Query Match 5.2%; Score 106.4; DB 3; Length 2150;

Best Local Similarity 51.2%; Pred. No. 5.9e-21;

Matches 310; Conservative 0; Mismatches 281; Indels 15; Gaps 2;

QY 301 CGAGCGCATGCACTGCTGCTGCTCTCTCTGCGCTGCGCTCTCTCTTTTGGGGGCA 360

DB 814 CAAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 873

QY 361 GCTTTTGGGAGCAGCAGATGTTTCTACCTGATGAGAGCCGCTGCGAGCGATGAT 420

DB 874 GCTCTTCAACGAGAACCTGAGCTGTTCTCTCTATGAGCGCTGAGCAGCTGAGCA 933

QY 421 GACCTTCAGAGAGCAGCAGCCTGATGCTGACATGAGCTGTCGGGATCTGATAGGGC 480

DB 934 AAAACTGTACCCCGGGAGCGCGCTTCCCTGACAGGGGAGCGGGAGACATGCTGAGCGC 993

QY 481 CGCTTTCTTGGCAGATGAGCGCTCTTTGATGCTTACATGAGAACTGTCCTCCCGAGACA 540

DB 994 TCTTACCGCTGCGATCTTTTGGGTTTCCAGCTATAGCCCCGAGAGCAGTGGGGGCG 1053

QY 541 GTCC-----AGCCTTTTCAGTGGGAGAACAGCGCGGCGCTGTGTTCTGCACC 588

DB 1054 CAACCTCACCACTGTGGGCACTTTGGGGGAGCCACTAACAGGTGATGCTCTCTCGCC 1113

QY 589 TGGCTGTGACATCATCCCAAGATGAATCATCCCGGGGCTACAGCAGGCTCC---T 645

DB 1114 ACTGTGCTCTCTACCGCAGAGAGTGTGAGCTGTGTGAGCAGCAGCGCTGTGCAAAA 1173

QY 646 GTGAGTCAACAGCCCTTTGAGTGTGAGAGAGAGCGCTGCTCTACAGCCAGTGTGT 705

DB 1174 GTGGCCACTCAACGCTGGAGAGCTTCGAGAGAGAGTGTGCAAGTACCGCAGGTGTGT 1233

QY 706 GCTCAAGAGAGTGGGCTTTTCAACCTGCACTGCTCTACCGGCTGTGTAAGACCCCTC 765

DB 1234 TATCAAGGCGTGGGCTGTGATGTGCTGTGTGGCCGCGCTTAAAGATCCAGC 1293

QY 766 CCTCAACCTGATATGTCGACCTGTGCTGGGAGCCCGGCGCTGTTCGTTCCGAGA 825

Db 1294 CTGGAAGTCAAGTCAATCCACTAGTATGATCTCTGCTGTGACAGCTCCGCAT 1353

QY 826 AGCACAAGAGGAGATCTCATGATGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 885

Db 1354 CCGCTGCGCTCAGCGGCTCTATCCGGGAAAGCTACAGGTGTGCGAAGCGGGATCCAG 1413

QY 886 ACTCAA 891

Db 1414 AGCCCA 1419

RESULT 9

US-09-471-867-1

; Sequence 1, Application US/09471867

; Patent No. 6455289

; GENERAL INFORMATION:

; APPLICANT: Uchimura, Kenji

; APPLICANT: Muramatsu, Hideki

; APPLICANT: Kadomatsu, Kenji

; APPLICANT: Kanagai, Reiji

; APPLICANT: Habuchi, Osami

; APPLICANT: Muramatsu, Takashi

; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND

; FILE REFERENCE: TOYAMA1.001AUS

; CURRENT APPLICATION NUMBER: US/09/471,867

; CURRENT FILING DATE: 1999-12-23

; EARLIER FILING DATE: 1999-03-05

; EARLIER APPLICATION NUMBER: JP 10-54007

; EARLIER FILING DATE: 1998-03-05

; EARLIER APPLICATION NUMBER: JP 10-177844

; EARLIER FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 2150

; TYPE: DNA

; ORGANISM: Mus musculus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (470)...(1918)

US-09-471-867-1

Query Match 5.2%; Score 106.4; DB 4; Length 2150;

Best Local Similarity 51.2%; Pred. No. 5.9e-21;

Matches 310; Conservative 0; Mismatches 281; Indels 15; Gaps 2;

QY 301 CGAGCGCATGCACTGCTGCTGCTCTCTCTGCGCTGCGCTCTCTTTTGGGGGCA 360

DB 814 CAAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 873

QY 361 GCTTTTGGGAGCAGCAGATGTTTCTACCTGATGAGAGCCGCTGCGAGCGATGAT 420

DB 874 GCTCTTCAACGAGAACCTGAGCTGTTCTCTCTATGAGCGCTGAGCAGCTGAGCA 933

QY 421 GACCTTCAGAGAGCAGCAGCCTGATGCTGACATGAGCTGTCGGGATCTGATAGGGC 480

DB 934 AAAACTGTACCCCGGGAGCGCGCTTCCCTGACAGGGGAGCGGGAGACATGCTGAGCGC 993

QY 481 CGCTTTCTTGGCAGATGAGCGCTCTTTGATGCTTACATGAGAACTGTCCTCCCGAGACA 540

DB 994 TCTTACCGCTGAGATCTTGGTTCAGAGCTGATACCCCGGAGCAGTGGGGGCG 1053

QY 541 GTCC-----AGCCTTTTCAGTGGGAGAGAGAGCGCGGCGCTGTGTTCTGCACC 588

DB 1054 CAACCTCACCACTGTGGGCACTTTGGGGAGCCACTAACAGGTGATGCTCTCTCGCC 1113

QY 589 TGGCTGTGACATCATCCCAAGATGAATCATCCCGGGGCTCACTGAGGCTCC---T 645

DB 1114 ACTGTGCTCTCTACCGCAGAGAGTGTGAGCTGTGTGAGCAGCAGCGCTGTGCAAAA 1173

QY 646 GTGAGTCAACAGCCCTTTGAGTGTGAGAGAGGCTCTCCGCTCTTACAGCAGCTGTGT 705

Db 488 GACCTCCTGGGAGACCTCTAGACTGCGACCTCTACTTCTTGAGAACTCATCAAGCCG 547
QY 527 GGTCCCGGAGAGAGTCCAGCC-----TCTTTCAGTGGGAGAGACAGCGCCCTGTGT 580
Db 548 CCGCCGGTCAACACACACACCGAGATCTTCCGCGGGGCGACCGGGTCTCTGCG 607
QY 581 TCTGCACTGCTGTGACATCATCCCAAGAAATCAATC--CCCGGCTCTACTGC 637
Db 608 TCCCGGCTGTGTGACACCTCCGCGGCGCAGCGACCTGTGCTCTGAGAGAGGGGACTGT 667
QY 638 AGGCTCCTGTGCAATCAAGACCCCTTGTAGTGTGAGAGAGCGCTGCGCTCTACAG 697
Db 668 GTGCGCAAGTGGGCTACTCAACCTTACCTGCGCGCGGCGGCGGCGAGCGAGC 727
QY 698 CACGTGTGCTCAAGAGAGTGGCTTCTTCACTGACGTCTCTACCCGCTGTGAAA 757
Db 728 CACGTGTGCTCAAGAGAGTGGCTTCTTCACTGACGTCTCTACCCGCTGTGAAA 787
QY 758 GACCCCTTCTCAACCTTCTGATGTGCTGCTGCGGAGCCCGGGGCTGTTCGT 817
Db 788 GACCCGGGATTAACCTCAAGGTCTACCTGCTGCGAGCCCGCGGCTATCTGGCT 847
QY 818 TCCCGAGAGCGCAAGAGGAGATCTGATGATGACAGTGCATTTGATGGGCGAGCAT 877
Db 848 TCGGCGAGCGAGACCTTCCGCGAGACGTACCGGCTCTGGGCGCTCTGTACGGCAGCGG 907
QY 878 GAGCAGAACTCAAGAGAGAGAGACCACTTACTGATGATGAGTCACTGTCCCAAGC 937
Db 908 AGGAACCCCTCAACCTGACGTGACGACCTGACCGAGTGTGCGAGGACTCTCCAC 967
QY 938 CAGCTGAGATCTCAAGACCATCTCTGCTGCGCAAGGCGCTGCGAGAGCGTCTCTG 997
Db 968 TCGGTGTCCACCGGCTCTCATGCGGCGCGCGGCTCAAGGGCAG-----TTCATG 1018
QY 998 CTTGTGCGTATGAGAGACTGTGCTGAGCCCTGTGCGCCGAGCTTCCGAGATGTATGA 1057
Db 1019 TTGTGGGCTACAGAGACGTGGCTCGGAGCCCTATGAAGAACCGGAGATCTACGGG 1078
QY 1058 TTCGTGGGATTTGATCTTCTCCCATCTTCAAGCTGGGTGATATACATCACCAGCG 1117
Db 1079 TTCCTGGGATCTCCGCTGAGAGACGACGTGGCGCGCTGATCCAGAACACAGCGGGG 1138
QY 1118 AAGGCGATGGGTGACACGCTTTCACACAAATGCCAGGATCCCTTAATGTCTCCAG 1177
Db 1139 GACCCCAACCCCTGGGCGAGCAACAAATACGGACCGTGGGAGAACTGGGGCGCGAG 1198
QY 1178 GCTTGGCGCTGTGCTTGTCCCTATGAAGGTTTCTGACTTCAAGAGCCTGTGGCGAT 1237
Db 1199 AAGTGGGCTTCCGCTCTCTCTACGATGCTGTGCGCTTTCGCGAGAGCGCTCCAGCAG 1258
QY 1238 GCCATGATTTGCTGGGCTACCGCGCATGATCTGAACAGAGAGAGAGAGCTGTG 1297
Db 1259 GTGCTGGCGGAGCTGCGCTCTCTCTACGATGCTGTGCGCTTTCGCGAGAGCGCTCC 1318
QY 1298 CTGATCTTCTG 1309
Db 1319 GTGAGCTGTG 1330

RESULT 5
US-08-655-878-1
Sequence 1, Application US/08655878
Patent No. 5827713

GENERAL INFORMATION:
APPLICANT: FUKUTA, MASAKAZU
APPLICANT: HAHUCHI, OSAMI
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET:
CITY:

STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentId
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2354
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Chick
TISSUE TYPE: Embryo chondrocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 211..1584
FEATURE:
IDENTIFICATION METHOD: P
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 211..309
FEATURE:
IDENTIFICATION METHOD: P
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 310..1584
FEATURE:
IDENTIFICATION METHOD: P
FEATURE:
NAME/KEY: transmembrane domain
LOCATION: 280..321
FEATURE:
IDENTIFICATION METHOD: P
FEATURE:
NAME/KEY: potential N-glycosylation site
LOCATION: 394..402
FEATURE:
IDENTIFICATION METHOD: S
FEATURE:
NAME/KEY: potential N-glycosylation site
LOCATION: 427..435
FEATURE:
IDENTIFICATION METHOD: S
FEATURE:
NAME/KEY: potential N-glycosylation site
LOCATION: 493..501
FEATURE:
IDENTIFICATION METHOD: S
FEATURE:
NAME/KEY: potential N-glycosylation site
LOCATION: 916..924
FEATURE:
IDENTIFICATION METHOD: S
FEATURE:
NAME/KEY: potential N-glycosylation site
LOCATION: 1405..1413
FEATURE:
IDENTIFICATION METHOD: S
FEATURE:
NAME/KEY: potential N-glycosylation site
LOCATION: 1537..1545
FEATURE:
IDENTIFICATION METHOD: S

US-08-655-878-1

QY 121 CAGCTTCCGAGGAGAGTGCCTTCTCAGGCCGCTTGCAGAGTCTTCCACTTCAGAC 180
DB 121 CAGCTTCTGGAGAGAGTGCCTTCTCAGGCCGCTTGCAGAGTCTTCCACTTCAGAC 180
QY 181 AATGCTACTGGCTAAATAAATGAAGCTCCTGCTTTCTGTTTCCAGATGGCCATCTT 240
DB 181 AATGCTACTGGCTAAATAAATGAAGCTCCTGCTTTCTGTTTCCAGATGGCCATCTT 240
QY 241 GGCTTATTTCTCCAGATGTACAGCCACACATCAGCTCCTGCTGTATGAAGGACAGCC 300
DB 241 GGCTTATTTCTCCAGATGTACAGCCACACATCAGCTCCTGCTGTATGAAGGACAGCC 300
QY 301 CGAGCGCATGACAGTGTGTTCTGTTCTTCCGAGCTGCTGCTCTTCTTTTGGGAGCA 360
DB 301 CGAGCGCATGACAGTGTGTTCTGTTCTTCCGAGCTGCTGCTCTTCTTTTGGGAGCA 360
QY 361 GCTTTTGGGAGACACCCAGATGTTTCTACCTGATGAGCCCGCTGGCAGCTGTGGAT 420
DB 361 GCTTTTGGGAGACACCCAGATGTTTCTACCTGATGAGCCCGCTGGCAGCTGTGGAT 420
QY 421 GACCTTCAAGCAGACACCGCTGGATGCTGACATGGCTGGGGATGCTGATACGGGC 480
DB 421 GACCTTCAAGCAGACACCGCTGGATGCTGACATGGCTGGGGATGCTGATACGGGC 480
QY 481 GCTCTTCTTGGCAGATGAGCTGTTTGAATGCTTACATGGAACCTGTCCTCCGAGACA 540
DB 481 GCTCTTCTTGGCAGATGAGCTGTTTGAATGCTTACATGGAACCTGTCCTCCGAGACA 540
QY 541 GTCGAGCTCTTCAAGTGGAGAACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 541 GTCGAGCTCTTCAAGTGGAGAACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 601 CATCCCAAGATGAATCATCCCGCGGCTCAGTGCAGAGGCTGCTGCTGCTGCTGCTGCT 660
DB 601 CATCCCAAGATGAATCATCCCGCGGCTCAGTGCAGAGGCTGCTGCTGCTGCTGCTGCT 660
QY 661 CTTTGAAGTGGAGAGAGGCTGCTGCTTCTACAGCCAGCTGCTGCTGCTGCTGCTGCT 720
DB 661 CTTTGAAGTGGAGAGAGGCTGCTGCTTCTACAGCCAGCTGCTGCTGCTGCTGCTGCT 720
QY 721 CTTTCTCAACCTGAGTGCCTTACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 721 CTTTCTCAACCTGAGTGCCTTACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 781 CGTGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 781 CGTGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 841 TCTCATGATGACAGTGCATTTGATGGGAGCAGATGAGCAGAACTCAAGAGAGAGA 900
DB 841 TCTCATGATGACAGTGCATTTGATGGGAGCAGATGAGCAGAACTCAAGAGAGAGA 900
QY 901 CCAACCCCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB 901 CCAACCCCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 CCAAGTCTTGGCCCAAGGCTGCAAGAGAGCTTACTGCTTGTGGCTATGAGGAGCTGGC 1020
DB 961 CCAAGTCTTGGCCCAAGGCTGCAAGAGAGCTTACTGCTTGTGGCTATGAGGAGCTGGC 1020
QY 1021 TCGAGCCCTTGGCCCAAGGCTTCCGATGATGAATTCGTGGATGATGATGATGATGAT 1080
DB 1021 TCGAGCCCTTGGCCCAAGGCTTCCGATGATGAATTCGTGGATGATGATGATGATGAT 1080
QY 1081 CCAATTTAGACCTGGGTGATTAACATACCCGAGGCAAGGGCATGGGTGACCACTGTT 1140
DB 1081 CCAATTTAGACCTGGGTGATTAACATACCCGAGGCAAGGGCATGGGTGACCACTGTT 1140
QY 1141 CCACGCAATGCGAGGAGTGCCTTAATGATGCTCCAGAGCTGCGCTGCTGCTGCTGCT 1200
DB 1141 CCACGCAATGCGAGGAGTGCCTTAATGATGCTCCAGAGCTGCGCTGCTGCTGCTGCT 1200
QY 1201 TGAAGAGTTCCTGCACTTGAAGAGCTGTGGGATGCCATGAATTTGCTGGGCTACCG 1260

DB 1201 TGAAGAGTTCCTGCACTTGAAGAGCTGTGGGATGCCATGAATTTGCTGGGCTACCG 1260
QY 1261 CCACGTCAGATCTGAACACAGAGAAACCTGTTGCTGGATCTTCTGCTTACCTGGAC 1320
DB 1261 CCACGTCAGATCTGAACACAGAGAAACCTGTTGCTGGATCTTCTGCTTACCTGGAC 1320
QY 1321 TGTCCCTGAGCAATGCATCAAGAGGTTGAGAGGCTTGTGCTGCCACCTGGGTGAGCC 1380
DB 1321 TGTCCCTGAGCAATGCATCAAGAGGTTGAGAGGCTTGTGCTGCCACCTGGGTGAGCC 1380
QY 1381 TCACTCACTTCTCTGAATGCTTCTGAGCTTGCCTACATCTCTGAGCCTTAACATCAG 1440
DB 1381 TCACTCACTTCTCTGAATGCTTCTGAGCCTTGCCTACATCTCTGAGCCTTAACATCAG 1440
QY 1441 TCTGTGGGTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB 1441 TCTGTGGGTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1501 GTGTCCATGCTTGTCTGTAAGAAACAGACTGGGAGACCTTATGTAGCAGCATCCAC 1560
DB 1501 GTGTCCATGCTTGTGTGCTGTAAGAAACAGACTGGGAGACCTTATGTAGCAGCATCCAC 1560
QY 1561 CAGTGAACAGGATGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1620
DB 1561 CAGTGAACAGGATGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1620
QY 1621 AGACTTGTGGCTGGAGGCTATTAGACAGCAGATCAGTACAGTGAATGATCCATA 1680
DB 1621 AGACTTGTGGCTGGAGGCTATTAGACAGCAGATCAGTACAGTGAATGATCCATA 1680
QY 1681 ACCTCCCTGTCACATCTTGGCCATGAGGATGATCTTTTCCACCAAGAGCTCACAC 1740
DB 1681 ACCTCCCTGTCACATCTTGGCCATGAGGATGATCTTTTCCACCAAGAGCTCACAC 1740
QY 1741 ATTTCACAGAGATGAATCTGAGGCTTGGAGCTTCCAGTGGATGATCAAGAGAGAA 1800
DB 1741 ATTTCACAGAGATGAATCTGAGGCTTGGAGCTTCCAGTGGATGATCAAGAGAGAA 1800
QY 1801 GTGGGAACAGGTTGATGCTACTTATGAGCTTGCATCAGCAGCTATCGTAAATCAGA 1860
DB 1801 GTGGGAACAGGTTGATGCTACTTATGAGCTTGCATCAGCAGCTATCGTAAATCAGA 1860
QY 1861 AATATGAACAAATCTCTGACAAAGAGCAAGCTCTTAAGTTCACAGAGTGCCTGGGC 1920
DB 1861 AATATGAACAAATCTCTGACAAAGAGCAAGCTCTTAAGTTCACAGAGTGCCTGGGC 1920
QY 1921 TGCATTTGAATATCACTTCCCTCTGCTGATTTTCCATCAGATGAGAGCTTGGACTGTG 1980
DB 1921 TGCATTTGAATATCACTTCCCTCTGCTGATTTTCCATCAGATGAGAGCTTGGACTGTG 1980
QY 1981 AAGCTGCCATCTGTATACTAAATCCCAATTAAGAAAAA 2032
DB 1981 AAGCTGCCATCTGTATACTAAATCCCAATTAAGAAAAA 2032

RESULT 3
US-08-899-514-1
; Sequence 1, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; APPLICANT: FUKUDA, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US

Db 361 GCTTTTGGGAGACACCCAGATGTTTTCTACGTATGAGCCCGCTGGCAGCTGTGAT 420
QY 421 GACCTTCAAGCAGACACCGCTGGATGCTGCATATGGCTGGGGATGCTGATACGGGC 480
Db 421 GACCTTCAAGCAGACACCGCTGGATGCTGCATATGGCTGGGGATGCTGATACGGGC 480
QY 481 GCTCTTCTTGTGCGACATGAGCGCTTTTGATGCCATATGAGACCTGGTCCCGAGACA 540
Db 481 GCTCTTCTTGTGCGACATGAGCGCTTTTGATGCCATATGAGACCTGGTCCCGAGACA 540
QY 541 GTCGAGCCTCTTTCAGTGGGAGAACAGCCGGCCCTGTGTTCTGCACCTGCTGTGACAT 600
Db 541 GTCGAGCCTCTTTCAGTGGGAGAACAGCCGGCCCTGTGTTCTGCACCTGCTGTGACAT 600
QY 601 CATCCACAGATGATAATCATCCCCGGGCTCAGTGCAGGCTCCGTGACATCAACAGCC 660
Db 601 CATCCACAGATGATAATCATCCCCGGGCTCAGTGCAGGCTCCGTGACATCAACAGCC 660
QY 661 CTTTGAGGTGGTGGAGAGGCGCTGCGCTCTACAGCCAGCTGTGCTCAAGAGGTGCG 720
Db 661 CTTTGAGGTGGTGGAGAGGCGCTGCGCTCTACAGCCAGCTGTGCTCAAGAGGTGCG 720
QY 721 CTTTCTACCTGAGTGCCTCTACCCGCTCTGTAAGACCCCTCCCTCAACCTGCATAT 780
Db 721 CTTTCTACCTGAGTGCCTCTACCCGCTCTGTAAGACCCCTCCCTCAACCTGCATAT 780
QY 781 CGTGACCTGCTCGGGAGACCCCGGGCGCTGTCCGTTCCCGAGAGCGCACAAAGGAGA 840
Db 781 CGTGACCTGCTCGGGAGACCCCGGGCGCTGTCCGTTCCCGAGAGCGCACAAAGGAGA 840
QY 841 TCTCATGTTATACAGTCCATTTGATGAGGGGACAGATGAGAGAACTCAAGAGAGAGA 900
Db 841 TCTCATGTTATACAGTCCATTTGATGAGGGGACAGATGAGAGAACTCAAGAGAGAGA 900
QY 901 CCAACCTCTACTATGTATGATGACAGTATCTGCCAAGCCAGCTGAGATCTACAAAGACAT 960
Db 901 CCAACCTCTACTATGTATGATGACAGTATCTGCCAAGCCAGCTGAGATCTACAAAGACAT 960
QY 961 CCAAGTCTTGGCCAAAGGCGCTGCGAGAAAGCTACTGCTTGGCGCTATAGAGACCTGGC 1020
Db 961 CCAAGTCTTGGCCAAAGGCGCTGCGAGAAAGCTACTGCTTGGCGCTATAGAGACCTGGC 1020
QY 1021 TCGAGCCCTGTGGCCAGACATTCGCAATGTATGAATGCTGGGATGGAATCTTGCC 1080
Db 1021 TCGAGCCCTGTGGCCAGACATTCGCAATGTATGAATGCTGGGATGGAATCTTGCC 1080
QY 1081 CCACTTCTCAGACCTGGGTGCTATACATCACCCGAGGCAAGGGCATGGGTGACACGCTTT 1140
Db 1081 CCACTTCTCAGACCTGGGTGCTATACATCACCCGAGGCAAGGGCATGGGTGACACGCTTT 1140
QY 1141 CCACACAATGCCAGGGATGCCCTTAATGTCTCCAGGCTTGGGGCTGTGCTTGGCCCTA 1200
Db 1141 CCACACAATGCCAGGGATGCCCTTAATGTCTCCAGGCTTGGGGCTGTGCTTGGCCCTA 1200
QY 1201 TGAAGAGGTTCTGACTTCAAGAAAGCCTGTGGGATGCCATGATTTGCTGGGCTACCG 1260
Db 1201 TGAAGAGGTTCTGACTTCAAGAAAGCCTGTGGGATGCCATGATTTGCTGGGCTACCG 1260
QY 1261 CCACGTGAGATCTGAACAGAGACAGAAAGCTTTGCTGAGATCTGTCTACCTGAGAC 1320
Db 1261 CCACGTGAGATCTGAACAGAGACAGAAAGCTTTGCTGAGATCTGTCTACCTGAGAC 1320
QY 1321 TGTCCCTGAGCAATTCACACTAGAGGGGTGAGAAAGCTTTGCTGCCACCTGATGTGAGCC 1380
Db 1321 TGTCCCTGAGCAATTCACACTAGAGGGGTGAGAAAGCTTTGCTGCCACCTGATGTGAGCC 1380
QY 1381 TCACTCACTTCTCTGATGATGCTTGGAGCTTGGCTGACATCTCTGAGCCCTTAACATG 1440
Db 1381 TCACTCACTTCTCTGATGATGCTTGGAGCTTGGCTGACATCTCTGAGCCCTTAACATG 1440
QY 1441 TCTGTGGGTATCACACTGAGTGTGAGTGTGTGCACACGCTGCTCAAGAGAGACTTTT 1500
Db 1441 TCTGTGGGTATCACACTGAGTGTGAGTGTGTGCACACGCTGCTCAAGAGAGACTTTT 1500

QY 1501 GTGTCATGCTGTGCTGCTAGAAAAACAGACTGGGGAACCTTATGTGACGACATCCAC 1560
Db 1501 GTGTCATGCTGTGCTGCTAGAAAAACAGACTGGGGAACCTTATGTGACGACATCCAC 1560
QY 1561 CAGTGAACAGAGGATATTGCTCTTCTTCTTCTTCTGATCTTCTGCTGGGAGACTTCAG 1620
Db 1561 CAGTGAACAGAGGATATTGCTCTTCTTCTTCTTCTGATCTTCTGCTGGGAGACTTCAG 1620
QY 1621 AGACTTGTGGCTGGAGGCTTATTAAGCAGCAGACATGATGATGATGATCCATTA 1680
Db 1621 AGACTTGTGGCTGGAGGCTTATTAAGCAGCAGACATGATGATGATGATCCATTA 1680
QY 1681 ACCCTCCTGTCACATCTTCCCAATGGGGAATGATCTTTCACCAAGAGCTCACAG 1740
Db 1681 ACCCTCCTGTCACATCTTCCCAATGGGGAATGATCTTTCACCAAGAGCTCACAG 1740
QY 1741 ATTTCCACAGAGATGCAAAATTTCTGAGCCCTTGGAGTTCCCACTGATTCAGAGAGAA 1800
Db 1741 ATTTCCACAGAGATGCAAAATTTCTGAGCCCTTGGAGTTCCCACTGATTCAGAGAGAA 1800
QY 1801 GTGGGAACAAGTGGATGCTACTATGAGCTTGAACATCAGACTATCGGTAATCAGA 1860
Db 1801 GTGGGAACAAGTGGATGCTACTATGAGCTTGAACATCAGACTATCGGTAATCAGA 1860
QY 1861 AATATGAACAAATCTCTGACAAAAAGACAGCTTTAGTTTCACAGGGTCCCTGGGC 1920
Db 1861 AATATGAACAAATCTCTGACAAAAAGACAGCTTTAGTTTCACAGGGTCCCTGGGC 1920
QY 1921 TGCATTTGAATATCAGCTTCCCTCTGCTGATTTTCCCATCATAGAAAGACTTGG 1980
Db 1921 TGCATTTGAATATCAGCTTCCCTCTGCTGATTTTCCCATCATAGAAAGACTTGG 1980
QY 1981 AAGCTGCATCTGTATATCTAAATTCACAAATAGAAAAA 2032
Db 1981 AAGCTGCATCTGTATATCTAAATTCACAAATAGAAAAA 2032

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOXYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 100.0%; Score 2032; DB 4; Length 2032;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGCTCAGGCGAGATGCTTCCAGTGTGGGGAAATGCTTCTCATTTGCTTCCAG 60
Db 1 GGCTCAGGCGAGAGGCTTCCAGTGTGGGGAAATGCTTCTCATTTGCTTCCAG 60
QY 61 CCCACTCAAGAGATCTCCCAACCCCTTGAAGTCTGAGAGTGTAAAGCTGTACTTTCA 120
Db 61 CCCACTCAAGAGATCTCCCAACCCCTTGAAGTCTGAGAGTGTAAAGCTGTACTTTCA 120

R. Emmett-Buck, M.D., Ph.D.
CDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.illnl.gov/dbip/image/image.html

Search completed: January 10, 2003, 21:29:53
Job time : 2924 secs

Db 240 GAAGATCATCTGCAAAAGCCAGGTGACATAGTCAAGCCATCCAAACCCCTCCCTGAAGC 299

QY 979 CCTGCAGGAGACGCTACTCTGTTGCGCTATAGAGACCTGGCTCGAGCCCTGTGGCCCA 1038

Db 300 TCTGCGAGCAGCGCTACCTGTTCTGTAGATAGAGACCTGGCTGGGACACCCCTGGCCCA 359

QY 1039 GACTCCCGAGATATGAATTCGTGGGATTTGAATTCCTTGGCCCATCTTCAGACCTGGGT 1098

Db 360 GAGGACAGAGCTATTAATTT-CTGGGGTTGGATTTTGGCCCATCTTCAGACATGAGGT 418

QY 1099 GCATTAACATACCCGAGGAGGCGATGGGTGACACGCTTTCACACAAATGCCAGGA 1158

Db 419 TCACATATGTCACCCGAGGAGGAGGATGGGTGACAGCTTCACATACAGCCAGGA 478

QY 1159 TGGCCTTAATGTTCCAGGCTTGGCGCTGGTGTCTTGGCCCTA-TGAAGGTTTTCGAC 1217

Db 479 CGCCCTCAAGCTCTTCACAGCGGTGGCTGTGCTTACTTACCGAAAGGTTTCCGAC 538

QY 1218 TTCAAGAAAGCTGTGGCGATGCCAT-GAATTTGCTGGGCTACCGCCAGCTCAGATCTGAA 1276

Db 539 TTCAAGATGCTCGGTGAGGCTATGATTTTGTGGATACCTCCAGGTCAAGATCTCAA 598

QY 1277 CAGAAGCAGAGAACCTGTGGTGGATCTTGTCTTACTGAGCTGGCTCCCTGAGCAATC 1336

Db 599 CAGAGACA-AGGCACTGTCCCTGCGATCTTCTCTCTCCCATATCTTGGGGCAGCTC 657

QY 1337 CACTAAG-AGGTTGAGAAAGGCTTTGCTGCACACCTGTGTGACCC 1380

Db 658 TTCCGAGCAGCGTTAAGAGAGGCTGTGACACCCCTGTGTCAGGCC 702

RESULT 13
AM572390/c 419 bp mRNA linear EST 13-MAR-2000
LOCUS xu09a06.x2 NCI_CGAP_Col4 Homo sapiens cDNA clone IMAGE:2799634 3'
DEFINITION mRNA sequence.
ACCESSION AM572390
VERSION AM572390.1 GI:7237123
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 419)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/nclogap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov/image/html/resources.shtml>
Seq. primer: 40UP from Gibco
High quality sequence stop: 343.

FEATURES

source
1. 419
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2799634"
/clone_lib="NCI_CGAP_Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.7 kb. Life technologies catalog #: 11531-019"

BASE COUNT 122 a 78 c 89 g 130 t

ORIGIN

Query Match 16.4% Score 332.4; DB 10; Length 419;
Best local similarity 99.7% Pred. No. 1.4e-60;
Matches 333; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1684 TCCCGTCACATCTTGGCCCATGGGATGATCTTTCACCAAGAGCTCACCAGCAT 1743

Db 419 TCCCGTCACATCTTGGCCCATGGGATGATCTTTCACCAAGAGCTCACCAGCAT 360

QY 1744 TTCCACAGAGATGCATAATCTTGAGCCCTTGGAGTTCCAGTGATTCAGAGAGAGTG 1803

Db 359 TTCCACAGAGATGCATAATTTTGGAGCCCTTGGAGTTCCAGTGATTCAGAGAGAGTG 300

QY 1804 GGAACAGAGTGAGAGCTTACTATGAGCTTACCATCAGCATCAGTATGAGAAAT 1863

Db 259 GGAACAGAGTGAGAGCTTACTATGAGCTTACCATCAGCATCAGTATGAGAAAT 240

QY 1864 ATGAACAAATCTCTGACAAAGAGCAGCTCTTAAGTTCCAGAGGCTGCTGGCTGC 1923

Db 239 ATGAACAAATCTCTGACAAAGAGCAGCTCTTAAGTTCCAGAGGCTGCTGGCTGC 180

QY 1924 ATTGAATATCATCTCCCTCGCATTTTCCATCATATAGAGCTTGAAGTGAAG 1983

Db 179 ATTGAATATCATCTCCCTCGCATTTTCCATCATATAGAGCTTGAAGTGAAG 120

QY 1984 CTGCCATCTGTTAATCTAAATTTCCCAATTAAG 2017

Db 119 CTGCCATCTGTTAATCTAAATTTCCCAATTAAG 86

RESULT 14
B0947021 902 bp mRNA linear EST 21-AUG-2002
LOCUS B0947021
DEFINITION AGENCOURT 8951359 NCI_CGAP_CO24 Mus musculus cDNA clone
IMAGE:6474928 5', mRNA sequence.
ACCESSION B0947021
VERSION B0947021.1 GI:22362499
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scuriognath; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 902)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM14012 row: n column: 17
High quality sequence stop: 633.

FEATURES

source
1. 902
Location/Qualifiers
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6474928"
/clone_lib="NCI_CGAP_CO24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Technology. Note: this is a NCI-CGAP Library."

BASE COUNT 228 a 246 c 210 g 213 t

ORIGIN


```

/c1one_lib="NCI_CGAP_CO24"
/lab_host="DH10B (T1 phage-resistant)"

```

REFERENCE
1 Mammalia; Eutheria;
1 (bases 1 to 852)

REFERENCE
1 Mammalia; Eutheria;
1 (bases 1 to 852)

Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brownj@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 434.

FEATURES

source

Location/Qualifiers
1..571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:567655"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site:1; Not 1;
Site:2; Sal 1; Starting library constructed using
Superscript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dt priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennan, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an Ecot of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 109 a 109 c 143 g 150 t
ORIGIN

Query Match 28.0%; Score 569.4; DB 13; Length 571;
Best Local Similarity 99.8%; Pred. No. 4.1e-111;
Matches 570; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 88 TGAATCTCAGCAGTGTAAAGCTTTACTTTACAGCTTCTCGGAGCGAGTCTTCTC 147
DB 1 TGAATCTCAGCAGTGTAAAGCTTTACTTTACAGCTTCTCGGAGCGAGTCTTCTC 60
QY 148 AAGCCCGCTTGCAGAGCTTCCACTTCAGCAGCATCTCTGCTTAAATAATGAGCT 207
DB 61 AAGCCCGCTTGCAGAGCTTCCACTTCAGCAGCATCTCTGCTTAAATAATGAGCT 120
QY 208 CCGTCTGTTTGTGTTTCCAGATGCGCATCTTGGCTCTATTTCCACATGATGACCA 267
DB 121 CCGTCTGTTTGTGTTTCCAGATGCGCATCTTGGCTCTATTTCCACATGATGACCA 180
QY 268 CAACATGAGCTCCGCTGTATGAAGGACAGCCGAGGAGTGCAGCGGCTGTTCTGC 327
DB 181 CAACATGAGCTCCGCTGTATGAAGGACAGCCGAGGAGTGCAGCGGCTGTTCTGC 240
QY 328 TTCCGAGGCTCTGAGCTCTCTTTTGGGGCAGCTTTTGGGAGCAGCCAGATGTTTT 387
DB 241 TTCCGAGGCTCTGAGCTCTCTTTTGGGGCAGCTTTTGGGAGCAGCCAGATGTTTT 300
QY 388 CTACCTGATGAGCCCGCTGCGACGTGTGATGACCTCAAGAGAGCAGCCGCTGGAT 447
DB 301 CTACCTGATGAGCCCGCTGCGACGTGTGATGACCTCAAGAGAGCAGCCGCTGGAT 360
QY 448 GCTGACATGAGCTGTGCGGATCTGATACGGGCGCTTTCTTGTGCGACATGAGCGCTT 507
DB 361 GCTGACATGAGCTGTGCGGATCTGATACGGGCGCTTTCTTGTGCGACATGAGCGCTT 420
QY 508 TGATGCTTACATGAGCTGTGCGGAGAGAGCTCCAGCTTTTTCAGTGGAGAGACAG 567
DB 421 TGATGCTTACATGAGCTGTGCGGAGAGAGCTCCAGCTTTTTCAGTGGAGAGACAG 480
QY 568 CCGGCGCTGTGTTCTGACCTGCTGTGACATCATCCCAAGATGAATCATCCCGG 627

DB 481 CCGGCGCTGTGTTCTGACCTGCGCTGTGACATCATCCCAAGATGAATCATCCCGG 540
QY 628 GGCTCAGTCAGAGCTCTCTGTGAGTCAACAG 658
DB 541 GGCTCAGTCAGAGCTCTCTGTGAGTCAACAG 571

RESULT 6
BF878439 680 bp mRNA linear EST 17-JAN-2001
LOCUS MRO-ET0109-191100-002-h06 ET0109 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF878439
ACCESSION BF878439
VERSION BF878439.1 GI:12268569
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 680)

REFERENCE
AUTHORS
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR0&t2=MR0-ET0109-
191100-002-h06&t3=2000-11-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 641.
Location/Qualifiers
1..680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0109"
/dev_stage="Adult"

FEATURES
source
1..680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0109"
/dev_stage="Adult"

/note="Organ: Lung-tumor; Vector: puc18; Site:1: Smal;
Site:2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 168 a 164 c 163 g 184 t
ORIGIN

Query Match 26.8%; Score 544.2; DB 12; Length 680;
Best Local Similarity 95.5%; Pred. No. 9.1e-105;
Matches 592; Conservative 0; Mismatches 24; Indels 4; Gaps 3;

QY 1382 CAGTCACTTCTCTGATGCTTGTGAGCCCTTGCT-ACATCTGTGAGCCTTAATCAATG 1440
DB 17 CAGTCACTTCTCTGATGCTTGTGAGCCCTTGCT-ACATCTGTGAGCCTTAATCAATG 76
QY 1441 TCTGTGGTATACACT--GAGTGTAGTGTCTGTCCACAGCTGCTCAAGCAGAGACTT 1498
DB 77 TCTGTGGTATACACTCGAGTGTGAGTGTCTGTCCACAGCTGCTCAAGCAGAGACTT 136

QY 1676 CATAAACCCTCCCTGTCACATCTTCCCAATGGGGAATGATCTTTTACCAAAAGACTCA 1735
 DB 333 CATAAACCCTCCCTGTCACATCTTCCCAATGGGGAATGATCTTTTACCAAAAGACTCA 294
 QY 1736 CCAGCATTTTCACAGAGATGCAAAATTCAGCCCTTGAGCTTCCCAATGATTCAGCA 1795
 DB 293 CCAGCATTTTCACAGAGATGCAAAATTCAGCCCTTGAGCTTCCCAATGATTCAGCA 234
 QY 1796 AGGAAGTGGGAACAAGTTCGATGCTTATAGCTTGACATCAAGCTATGCTAA 1895
 DB 233 AGGAAGTGGGAACAAGTTCGATGCTTATAGCTTGACATCAAGCTATGCTAA 174
 QY 1896 TCAGAAATATGAACAAATCTCTGCAACAAAGCAAGCTTAAAGTCCAGGCTGCC 1915
 DB 173 TCAGAAATATGAACAAATCTCTGCAACAAAGCAAGCTTAAAGTCCAGGCTGCC 114
 QY 1916 TGGGCTGATTTGAATATACCTTCCCTGTCATTTCCCATCAGATGAGACTTGCAC 1975
 DB 113 TGGGCTGATTTGAATATACCTTCCCTGTCATTTCCCATCAGATGAGACTTGCAC 54
 QY 1976 CTGTAAAGCTGCCATCTCTTAACTAAATTCACCAATAGAAAAA 2028
 DB 53 CTGTAAAGCTGCCATCTCTTAACTAAATTCACCAATAGAAAAA 1

RESULT 4
 AL709927 668 bp mRNA linear EST 22-MAR-2002
 LOCUS AL709927

DEFINITION DKFZP68602364.f1 666 (synonym: h1cc3) Homo sapiens cDNA clone
 DKFZP68602364.5, mRNA sequence.
 AL709927

ACCESSION AL709927 GI:19693282
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann
 S.
 EST (Duesterhoeft, et al.)
 Unpublished (1999)
 CONTACT: Duesterhoeft A
 MFS
 Am Klopferstr. 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert.
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.
 This clone (DKFZP68602364) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Source

1..668
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZP68602364"
 /clone_lib="686 (synonym: h1cc3)"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pTRi-plex2; Site_1: SfiI; Site_2: SfiI;
 cDNA-collection"
 BASE COUNT 132 a 202 c 178 g 154 t 2 others
 ORIGIN

Query Match 28.6%; Score 581.6; DB 9; Length 668;
 Best Local Similarity 98.2%; Pred. No. 9,9e-114;
 Matches 596; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 148 AAGCCGCTCTTGAAGTCTTCCACTTCAGACAGATGCTACGCTTAAAAAATGAGCT 207
 DB 60 AAGCCGCAAGCCAGAGTCTTCCACTTCAGACAGATGCTACGCTTAAAAAATGAGCT 119
 QY 208 CCGCTGTTCTCTGTTTCCAGATGCGCACTTGGCTGCTATTTCTTCCATGAGCA 267
 DB 120 CCGCTGTTCTCTGTTTCCAGATGCGCACTTGGCTGCTATTTCTTCCATGAGCA 179
 QY 268 CAACATGAGTCCCTGCTATATGAGGACAGAGCCGAGGCAATGACAGTCTGTTCTGTC 327
 DB 180 CAACATGAGTCCCTGCTATATGAGGACAGAGCCGAGGCAATGACAGTCTGTTCTGTC 239
 QY 328 TTCTGGGCTCTGCGCTCTTCTTTGGGGGAGCTTTTGGGACACCGACAGATTTT 387
 DB 240 TTCTGGGCTCTGCGCTCTTCTTTTGGGGGAGCTTTTGGGACACCGACAGATTTT 299
 QY 388 CTACCTGATGAGCCCGGCTGACGCTGATGATGACCTTCAAGCAGACCGGCTGAT 447
 DB 300 CTACCTGATGAGCCCGGCTGACGCTGATGATGACCTTCAAGCAGACCGGCTGAT 359
 QY 448 GCTGCACATGAGCTGCTGCGGATGATGATGAGGCGGCTTCTTGTGACATGAGCTTT 507
 DB 360 GCTGCACATGAGCTGCTGCGGATGATGATGAGGCGGCTTCTTGTGACATGAGCTTT 419
 QY 508 TGATGCTACATGAGCACTGCTGCTGCGGAGACAGTCCAGCTTTCAGTGGAGAACAG 567
 DB 420 TGATGCTACATGAGCACTGCTGCTGCGGAGACAGTCCAGCTTTCAGTGGAGAACAG 479
 QY 568 CCGGCGCCCTGTGCTTGTGACACCTGCTGATCATCATCCCAAGATGAAATATCCCCG 627
 DB 480 CCGGCGCCCTGTGCTTGTGACACCTGCTGATCATCATCCCAAGATGAAATATCCCCG 539
 QY 628 -GGCTCAGTGCAGGCTGCTGCTGAGTCAACAGAGCCCTTGAAGTGGGGAGAGGCTGCC 686
 DB 540 NGGCTCAGTGCAGGCTGCTGCTGAGTCAACAGAGCCCTTGAAGTGGGGAGAGGCTGCC 599
 QY 687 GCTCTTACAGCAGCAGTGTGCTCAGAGAGGCTGCTTCTTCAACCTGACGCTCTTACC 746
 DB 600 GCTCTTACAGCAGCAGTGTGCTCAGAGAGGCTGCTTCTTCAACCTGACGCTCTTACC 659
 QY 747 CGCTGCTGA 755
 DB 660 CGCTGCTGA 668

RESULT 5
 BM129080 571 bp mRNA linear EST 12-MAR-2002
 LOCUS BM129080
 DEFINITION i1f17c04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 cDNA clone IMAGE:567665 5' similar to TR:0915R3 0915R3
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
 BM129080
 ACCESSION BM129080.1 GI:17123632
 VERSION
 KEYWORDS
 SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 571)
 Melton, D., Brown, J., Kently, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemshaj, I., Searce, M., Brestelli, J., Grawohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Bistano, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas,
 M., Gibbons, M., McCann, R., Cole, R., Tsagarelashvili, R., Williams, T.,
 Jackson, Y. and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Other ESTs: i1f17c04.x1
 TITLE
 JOURNAL
 COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138

```

Db      44  AACCCAGCCAGAGCTTCCACTTCAGACACATGCTACTGCTTAAATAAATGAAGCT 103
QY      208  CCTGCTGTTTCTGTTTCCAGATGGCCATCTTGCTTAATCTTCCACATGTACAGCA 267
Db      104  CCTGCTGTTTCTGTTTCCAGATGGCCATCTTGCTTAATCTTCCACATGTACAGCA 163
QY      268  CACATCATGCTCCCTGCTTGAAGCAGACGCCGAGCGATGACGCTGCTGCTGTC 327
Db      164  CACATCATGCTCCCTGCTTGAAGCAGACGCCGAGCGATGACGCTGCTGCTGTC 223
QY      328  TTCTGCGCTTGCTGCTTCTTTTGTGGGAGAGCTTTTGTGGAGACCCAGCATGTTT 387
Db      224  TTCTGCGCTTGCTGCTTCTTTTGTGGGAGAGCTTTTGTGGAGACCCAGCATGTTT 283
QY      388  CTACCTGATGAGACCCGCTGCGACCGTGTGATGACCTTAAACAGACGCCGCTGGAT 447
Db      284  CTACCTGATGAGACCCGCTGCGACCGTGTGATGACCTTAAACAGACGCCGCTGGAT 343
QY      448  GCTGCACATGGCTGTGGGGATCTGATACGGGCGCTTCTTGTGGACATGAGCTCTT 507
Db      344  GCTGCACATGGCTGTGGGGATCTGATACGGGCGCTTCTTGTGGACATGAGCTCTT 403
QY      508  TGATGCTACATGGAACCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 567
Db      404  TGATGCTACATGGAACCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 463
QY      568  CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 627
Db      464  CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 523
QY      628  GGTCTACATGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 687
Db      524  GGTCTACATGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 583
QY      688  CTCTACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 747
Db      584  CTCTACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 643
QY      748  GCTCTGTAAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
Db      644  GCTCTGTAAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
QY      807  CCGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 864
Db      704  CCGTGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 763
QY      865  GATGGGAGAGATGAGAGAAA--CTCAGAGAGAGAGAG--AACCTACTATGTATG-C 920
Db      764  GATGGGAGAGATGAGAGAAA--CTCAGAGAGAGAGAG--AACCTACTATGTATG-C 823
QY      921  AGTTCATCTGCCAAA--GCCAGCTGAGATTTAC--AAGACCATCTCACTCTTGGCCAA 977
Db      824  AGTTCATCTGCCAAA--GCCAGCTGAGATTTAC--AAGACCATCTCACTCTTGGCCAA 883
QY      978  CCGTGCAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1035
Db      884  CCGTGCAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 943
QY      1036  CCAGACTTCCC 1046
Db      944  CCAGACTTCCC 954

```

RESULT 3
 LOCUS BM969292/2
 DEFINITION UI-CF-ENO-acp-1-21-0-UI-s1 UI-CF-ENO Homo sapiens cDNA clone
 ACCESSION BM969292
 VERSION BM969292.1 GI:19586879
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

593 bp mRNA linear EST 21-MAR-2002

```

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 593)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704447
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul.mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLY-A=yes.
FEATURES
source
Location/Qualifiers
1..593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-ENO-acp-1-21-0-UI"
/clone_id="UI-CF-ENO"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-ENO is a cDNA library containing the following
tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells.
The library was constructed according to Bonaldo, Lennon
and Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is CTGCTCAGGT.
TAG_LIB=UI-CF-ENO
TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_SEQ=CTGCTCAGGT"
BASE COUNT 157 a 127 c 137 g 172 t
ORIGIN
Query Match 29.1%; Score 591.4; DB 14; Length 593;
Best Local Similarity 99.8%; Pred. No. 8.3e-116;
Matches 592; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1436 ACATGCTGTGGGATACACACTGAGTGTGATGCTGCTACACAGCTCTAAGCAGAGA 1495
Db 593 ACATGCTGTGGGATACACACTGAGTGTGATGCTGCTACACAGCTCTAAGCAGAGA 534
QY 1496 CTTTGTGTCATGCTGTGCTAGAAACAGAGTGGGAGACCTTATGTAGACAGCAT 1555
Db 533 CTTTGTGTCATGCTGTGCTAGAAACAGAGTGGGAGACCTTATGTAGACAGCAT 474
QY 1556 CCACACAGTGAACAGAGGATTTCTCTCTTTTCTTGATCTTCTGCTGCTGAGAC 1615
Db 473 CCACACAGTGAACAGAGGATTTCTCTCTTTTCTTGATCTTCTGCTGCTGAGAC 414
QY 1616 TTGAGAGACTTTGGGCTGAGAGGCTTAAAGCAGACAGCACTATCAGTGAATGATC 1675
Db 413 TTGAGAGACTTTGGGCTGAGAGGCTTAAAGCAGACAGCACTATCAGTGAATGATC 354

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:55:29 ; Search time 2886 Seconds

(without alignments)
11403.059 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 1 ggcctgagagcagatgcct.....ataagaaaaaaaaaaaaa 2032

Sequence: 1 ggcctgagagcagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: IDENTITY_NUC
Gap10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estda.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_iny.*
20: em_gss_pln.*
21: em_gss_vtl.*
22: em_gss_fun.*
23: em_gss_man.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_too.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-----------------------------|
| 1 | 759.6 | 37.9 | 1923 | 11 | AK009113 Mus muscu |
| 2 | 751.4 | 37.5 | 1954 | 13 | B1823850 B1823850 603039012 |
| 3 | 551.4 | 29.1 | 593 | 14 | BM969292 UI-CF-ENO |
| 4 | 581.6 | 28.6 | 668 | 9 | AL709927 BKF2P6860 |
| 5 | 569.4 | 28.0 | 571 | 13 | BM129080 1f17c04.Y |
| 6 | 544.2 | 26.8 | 680 | 12 | Bf878439 MRO-ET010 |

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------------|
| 7 | 478 | 23.5 | 553 | 13 | BM128370 1f13d12.x |
| 8 | 478 | 23.5 | 553 | 13 | BM128831 1f17c04.x |
| 9 | 384.2 | 18.9 | 783 | 13 | BS963298 602827716 |
| 10 | 368.6 | 18.1 | 695 | 13 | BS964671 602831875 |
| 11 | 364.4 | 17.9 | 852 | 13 | BS966340 602832826 |
| 12 | 361.8 | 17.8 | 965 | 12 | BS579746 602095056 |
| 13 | 332.4 | 16.4 | 419 | 10 | AMS72390 X09a06.x |
| 14 | 324 | 15.9 | 902 | 14 | BS947021 AGENCOURT |
| 15 | 288 | 14.2 | 362 | 10 | AM002418 w61g03.x |
| 16 | 282.6 | 13.9 | 834 | 14 | BS084877 AGENCOURT |
| 17 | 275.2 | 13.5 | 358 | 12 | BF056840 7X10d12.x |
| 18 | 274.4 | 13.5 | 695 | 12 | BS857538 7G01a08.x |
| 19 | 263.6 | 13.0 | 735 | 12 | BS856652 7G01a09.x |
| 20 | 260.2 | 12.8 | 536 | 13 | BM245312 K0722507-x |
| 21 | 252.4 | 12.4 | 620 | 9 | A1824100 w146c01.x |
| 22 | 235.4 | 11.6 | 1067 | 17 | CNS03KN7 AL248380 Tetradon |
| 23 | 230.8 | 11.4 | 497 | 13 | BM246681 K0741E04- |
| 24 | 221 | 10.9 | 525 | 12 | BF197521 7o84a08.x |
| 25 | 192.6 | 9.5 | 849 | 17 | CNS04QFN AL302940 Tetradon |
| 26 | 182.4 | 9.0 | 417 | 10 | AMS72510 xq18g11.x |
| 27 | 167 | 8.2 | 568 | 9 | A1155260 u145c07.y |
| 28 | 158.8 | 7.8 | 657 | 17 | AG035205 Pan tTog1 |
| 29 | 156.2 | 7.7 | 517 | 12 | BF042384 BP250022A |
| 30 | 156.2 | 7.7 | 656 | 9 | AA522184 v198b09.r |
| 31 | 155.2 | 7.6 | 433 | 17 | A2405100 1N0173108 |
| 32 | 147.4 | 7.3 | 436 | 10 | AM081348 xc41b06.x |
| 33 | 143.4 | 7.1 | 429 | 10 | BS849113 BS849113 |
| 34 | 142.2 | 7.0 | 295 | 9 | A1614882 v198b09.y |
| 35 | 141.8 | 7.0 | 380 | 13 | BM435243 BM435243 1Om10F03 |
| 36 | 126.4 | 6.2 | 599 | 13 | BJ031352 BJ031352 |
| 37 | 123.8 | 6.1 | 2778 | 11 | BC017499 Homo sap1 |
| 38 | 120.6 | 5.9 | 961 | 14 | BO734733 AGENCOURT |
| 39 | 117.2 | 5.8 | 640 | 9 | AL681288 AL681288 |
| 40 | 114.2 | 5.6 | 634 | 9 | AU180328 AU180328 |
| 41 | 106.6 | 5.2 | 1169 | 14 | BS067314 AGENCOURT |
| 42 | 92.6 | 4.6 | 514 | 14 | BS065053 H4065A07- |
| 43 | 92.2 | 4.5 | 662 | 12 | BF344303 602017370 |
| 44 | 88 | 4.3 | 2070 | 11 | AK011202 Mus muscu |
| 45 | 86.6 | 4.3 | 322 | 9 | A1156825 u144c08.y |

ALIGNMENTS

RESULT 1
AK009113
LOCUS
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003618:carboxylate (Chondroitin 6/Keratan) sulfotransferase 4, full insert sequence.
ACCESSION AK009113
VERSION AK009113.1 GI:12843701
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_1lib:RIKEN full-length enriched mouse cDNA library
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
PUBMED 99279253
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE 2049374
PUBMED 11042159

This Page Blank (uspto)

| | |
|---------------------------|--|
| PT | New nucleic acid encoding cornea |
| PT | N-acetylglucosamine-6-sulfotransferase, useful for treatment, |
| PT | monitoring and diagnosis of macular corneal dystrophy . |
| XX | |
| XX | Claim 4; Fig 1A-D; 69pp; English. |
| CC | The present sequence encodes human corneal |
| CC | N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze |
| CC | sulfation of keratan sulfate (KS). Also described is a method for |
| CC | monitoring the effect of treatments for macular corneal dystrophy (MCD), |
| CC | and detecting susceptibility to MCD. (I) is located to chromosome 16q22, |
| CC | and has optimal immunological activity. (II) can be used to treat or prevent |
| CC | macular corneal dystrophy types I or II. (II) makes possible treatment |
| CC | of MCD without requiring keratoplasty or keratectomy. |
| XX | |
| SQ | Sequence 2544 BP; 460 A; 800 C; 733 G; 551 T; 0 other: |
| Query Match | 20.4%; Score 414.4; DB 24; Length 2544; |
| Best Local Similarity | 64.5%; Pred. No. 66-110; |
| Matches 654; Conservative | 0; Mismatches 351; Indels 9; Gaps 2 |
| OY | 302 GAAGCGATGACAGTGTCTTGTTGTCCTGGGCGCTGAGCTTCTTTTGGGGCAG 361 |
| Dd | |
| Dd | GGCGCGGTGCATGTGTCTGTCTGCTTCCTGTGGCGCTCGGCCCTCGNCTCTGGGCCAA 866 |
| OY | 362 CTTTTGGGACACCAACCATTGTTTCTTACCGTGATGAGACCCGCTGGCACGTGGAG 421 |
| Dd | |
| Dd | CTCTTACACACACACCCGACGCTTCTTAACATAAGAACCCGCGGGACACTGTGGACC 926 |
| OY | 422 ACCTTCAGACAGACACCGCTGATGTGACATAGCGCTGTCGGGATCTGATACGGCC 481 |
| Dd | |
| Dd | ACCGTGTCACAGGACAGCCCGCCAAACGCTGCACATGTGCTGCGGACACTGTGCGCTCC 986 |
| OY | 482 GTCTTCTGTGGACATGAGCGCTTTTATGTGCTTACATGAACTGTGTCCTGGAGACAG 541 |
| Dd | |
| Dd | GTCCTCTGTGGACATGGAATGGAAGCTGTGTTGATGCTCATCTG--CTTGGGCGCGACAAC 1043 |
| OY | 542 TCACACCTCTTCACTGGGAGAACAAGCGGGCGCTGTGTTGACACCTGCTATACATC 601 |
| Dd | |
| Dd | TCCGACCTCTTCCATGTGGGCGCTGAGACCTGTGACATGTGCTGCACACCGCCTGCAGTGCC 1103 |
| OY | 602 ATCCCACAAGATGAANTATCCCCGGGGCTACTGTCAGAGCTCTCTGTGCAATCAAGCCC 661 |
| Dd | |
| Dd | TTTTCCCGAGGCGCATCGACAGCAGAGCGCGTGTCAACAGCCACTGTGTGGCGGGCGAGTCC 1163 |
| OY | 662 TTGAGGTGTGAGAAGGCGTCCGCTCTACAGCCACAGTGTGCTCAAGAGAGTGGC 721 |
| Dd | |
| Dd | TTACACCTTGCGCCCGGAGAGCGTCCGCTCTACAGCCACAGTGTGCTCAAGAGAGTGGCC 1223 |
| OY | 722 TTCCTACACCTGCAATCCCTCTACCCGCTGCTGAAGAACCCTCCCTCAACCTGATATC 781 |
| Dd | |
| Dd | TTCTTCAACCTGCAAGTGTCTTACCCGCTGCTCAAGCAACCCCGCGCTCAACCTTAGCATTC 1283 |
| OY | 782 GTGACACGTGTCGGGAGACCCCGGGCGCTGTTCCTTCCCGAAGCACGCAAGAAGGAAT 841 |
| Dd | |
| Dd | GTGACCTGTGTGGCGGACCCCGGGCGCTGTGTGCGCTCCCGGAGACAGACAGCGAGCT 1343 |
| OY | 842 CTCATGATTGACAGTCGATGTGTATGGGGGACGATGAGCAAAACTAAGAAGAGAGAC 901 |
| Dd | |
| Dd | CTGGCGCTGTGCAACGSGATGTGTGTGGCAACCAAGGACAGTGGGTGAGAGCCGACACC 1403 |
| OY | 902 CAACCTCTACTATGTGATGCAGGTCTATCTGCCAAAAGCCAGCTGAGATGTACAAAGCATC 961 |
| Dd | |
| Dd | GAGCTGCGCGTGTGGAGAGTGTGGGCGGTGAGCCAGTACGATGCGGAGCGGCCACCA 1463 |
| OY | 962 CAGTCTTGGCCCAAGGCCCTGACAGACGCTAATCTGCTTGTGGCTATGAGAGACCTGGCT 1021 |
| Dd | |
| Dd | CTCAAAGCCGCAACCTTCTGTGCGGGCGGTACAGCGCCGTGGTGTGGAGAGACCTGGCG 1523 |
| OY | 1022 CGAGCCCTGTGGGCCCAAGACTTCCGAATGTATGATCTGAGTGATTTGGAATTTCTGGCC 1081 |
| Dd | |
| Dd | CGGAGCGCGTGTGGCAAAATCCGTGGCGTCTTACAGGCTTCAATGGGCTAAGTCTACAGCCA 1583 |


```

QY 422 ACCTTCAGAGAGACACCGCGCTGATGCTGCATATGAGCTGTGGGGATCTGATACGGCC 481
DB 1396 ACCCTGTGCGAGGCGACCGGCGCAACGCTGCATATGCGCTGGCGCTGATCTCT 1455
QY 482 GTCCTCTTGTGCGACATGACGCTGCTTGTGATGCTACATATGAACTGGTCCCGAGACAG 541
DB 1456 ATCTTTTGTGCGACATGAGAGCTGTGATGCTGCTACATATGAACTGGTCCCGAGACAG 1515
QY 542 TCCAGCCTCTTTCAGTGGAGAGACAGCGCGGCGCTGTGTTCTGCACCTGCTGTGACATC 601
DB 1516 TCCAGCCTCTTTCAGTGGAGAGACAGCGCGGCGCTGTGTTCTGCACCTGCTGTGACATC 1575
QY 602 ATCCCAAGATGAATCATATCCCGGGGCTCAGTCAAGGCTCCTGTGCTCAACAGCC 661
DB 1576 ATCCCAAGATGAATCATATCCCGGGGCTCAGTCAAGGCTCCTGTGCTCAACAGCC 1635
QY 662 TTGAGGTGTGAGAGAGCGCTGCGCTCTACAGCCAGTGTGTCTCAAGAGGTGCGC 721
DB 1636 TTGAGGTGTGAGAGAGCGCTGCGCTCTACAGCCAGTGTGTCTCAAGAGGTGCGC 1695
QY 722 TTCTTCAACCTGAGTCCCTCTACCGCTGTGTAAGACCCCTCCCTCAACCTGCTATATC 781
DB 1696 TTCTTCAACCTGAGTCCCTCTACCGCTGTGTAAGACCCCTCCCTCAACCTGCTATATC 1755
QY 782 GTGCACTGTGTCCGAGACCCCGGCGCTGTTCCTGTTCCGAGAACGCAAAAGGAGAT 841
DB 1756 GTGCACTGTGTCCGAGACCCCGGCGCTGTTCCTGTTCCGAGAACGCAAAAGGAGAT 1815
QY 842 CTCATGATGACATGCTGCTATGTATGGGAGACATATGCAAGAACTCAAGAGAGAGAC 901
DB 1816 CTCGACAGCGCAACGGATGTCTGTGGGCAACCAACGCAAGTGGGAGGCGGACCT 1875
QY 902 CACCCCTACTATGATGATGACAGTATGTGCCAAGCAGGAGTGAATCTCAAGACCATC 961
DB 1876 CACCTGCGCCGATTCGGAGAGTGTGCCGACGACGTCGGCCATGCGCGAGCGCCACA 1935
QY 962 CAGTCTTGTCCCAAGCGCTGCAGAGAACGCTACTGCTGTGGCCTATGAGAGACCTGCT 1021
DB 1936 CTCAGAGCGCGACCTCTCTCTGCGGCGGCGCTGACCGCTGTGGCCTGTGAGAGACCTGCG 1995
QY 1022 CGAGCCCTGTGGCGCACTTCCGATGTATGATTCGAGGATTTGGAATTTTCC 1081
DB 1996 CGGAGCGCGCTGCGAGAAATCCGGGCACTTACCGCTTACCGGCGCTGACCGCTACCGCA 2055
QY 1082 CATCTTCAGACCTGGGTGCATATACATCACCGGAGCAAGGSCATGGTGA-----CCAC 1135
DB 2056 CAGCTCGAGGCTGTGATCCACACATCACCCACGCGGTGCGGAGTCCGCAACCAATGAG 2115
QY 1136 GCTTCCACACAAATGCGAGGATGCGCTTATGTCCTCCAGGCTTGGCGCTGTGCTTTG 1195
DB 2116 GCTTCCACACTTGTGCTGCTGAAATGCGGCAAGCTTCCAGGCTTGGCGCTGACGCGTTG 2175
QY 1196 CCGATGAAAAGTTTCTGCACTTCAGAAAGCCTGTGGCATGCGCATGAAATTTGCTGGC 1255
DB 2176 CCGTCACTAAGATCTCGCGGTGAGAGAGTGTGCGCGCGCTGACACTGCTGGGC 2235
QY 1256 TACCCGACGTCAGATCTGAACAAAGAAAGAAACCTGTGCTGAGATCTCTG 1309
DB 2236 TACCGGCTGTGTACTGTGCGAGCACGAGTGTGACCTGTGAGTGTG 2289

```

RESULT 9
ABK54724
ID ABK54724 standard; cDNA: 517 BP.

XX
AC ABK54724;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human colon cancer-associated cDNA, SEQ ID No 194.
XX
KM Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.

```

XX OS Homo sapiens.
XX PN W0200212280-A2.
XX PD 14-FEB-2002.
XX PF 30-JUL-2001; 2001WO-US23826.
XX PR 03-AUG-2000; 2000US-223265P.
XX PR 02-OCT-2000; 2000US-237406P.
XX PR 20-MAR-2001; 2001US-277495P.
XX PR 03-JUL-2001; 2001US-302702P.
XX PA (CORI-) CORIXA CORP.
XX PI Pyle RA, Xu J, Secret H;
XX DR WPI; 2002-257462/30.
XX PT Novel polynucleotide encoding colon tumour polypeptides, useful as
XX PS vaccines for treating colon cancers
XX PS Claim 1, page 206; 425pp; English.
XX CC The invention relates to isolated polynucleotides (I) encoding colon
XX CC tumour polypeptides (II). (I) is useful for stimulating an immune
XX CC response in a patient and treating colon cancer in a patient.
XX CC Oligonucleotides derived from (I) are useful for determining the presence
XX CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
XX CC compositions, e.g. vaccines, and other compositions for the diagnosis
XX CC and treatment of colon cancer. A composition comprising a first component
XX CC selected from physiologically acceptable carriers and immunostimulants,
XX CC and an antigen-presenting cell expressing (II) is useful for inhibiting
XX CC development of cancer in a patient. (I) is useful in the design and
XX CC preparation of ribozyme molecules for inhibiting expression of tumour
XX CC polypeptides and (I), ABK54731-ABK55464 represent human colon cancer cDNA
XX CC sequences of the invention.
XX SO Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

Query Match 24.9%; Score 506.4; DB 24; Length 517;
Best Local Similarity 99.8%; Pred. No. 3.3e-137;
Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1352 GAAGGCTTTGCTCCACACCTGCTGCTGAGCTCACTTCTGTGATGCTTGTAGGCT 1411
DB 10 GAGGCGTTTGCTCCACCTGCTGCTGAGCTCACTTCTGTGATGCTTGTAGGCT 69
QY 1412 TGCTTACATCTGAGGCTTAATGATGCTGTGGTATACACAGAGTGTGAGTGTG 1471
DB 70 TGCTTACATCTGAGGCTTAATGATGCTGTGGTATACACAGAGTGTGAGTGTG 129
QY 1472 TCCACAGTGTCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTGTAGAAAACAGACTG 1531
DB 130 TCCACAGTGTCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTGTAGAAAACAGACTG 189
QY 1532 GGAAGCCTTGTGAGAGAGACATCCACAGAGAAAGAGGATGCTGCTTTT 1591
DB 190 GGAAGCCTTGTGAGAGAGACATCCACAGAGAAAGAGGATGCTGCTTTT 249
QY 1592 CTGATCTTCTGTGCGAGACTTCAGAGACTTGTGGCTGAGAGCCTATTAAGCAC 1651
DB 250 CTGATCTTCTGTGCGAGACTTCAGAGACTTGTGGCTGAGAGCCTATTAAGCAC 309
QY 1652 GACACAGTATCAGTGGATGATGCAATAAACCTCCCTGTCACATCTTGCCCAATGGGGA 1711
DB 310 GACACAGTATCAGTGGATGATGCAATAAACCTCCCTGTCACATCTTGCCCAATGGGGA 369
QY 1712 ATGATCTTTCACCAAGAGACTCACGAGATTTTCCACAGAGATGCAATTTGAGCCCT 1771
DB 370 ATGATCTTTCACCAAGAGACTCACGAGATTTTCCACAGAGATGCAATTTGAGCCCT 429

```


XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesized by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 39.68; Score 805.6; DB 22; Length 877;
Best Local Similarity 98.34; Pred. No. 2e-224;
Matches 834; Conservative 0; Mismatches 11; Indels 3; Gaps 2;

QY 114 ACTTTCACAGCTTCTGAGAGCGAGTCTCTTCAAGCCCGCTTGGCAAGCTTCCACT 173
DB 1 ACTTTCACAGCTTCTGAGAGCGAGTCTCTTCAAGCCCGCTTGGCAAGCTTCCACT 60
QY 174 TCACACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTTGGTTTCCAGATGG 233
DB 61 TCACACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTTGGTTTCCAGATGG 120
QY 234 CCATCTGGGCGCTATCTTCCACATGACAGCCCAACATCAGCTCCCTATATGAAG 233
DB 121 CCATCTGGGCGCTATCTTCCACATGACAGCCCAACATCAGCTCCCTATATGAAG 180
QY 294 CACAGCCGAGCGCATGAGTGTCTGCTTCTCTGCTGCGCTGAGCTCTTCTTTTG 353
DB 181 CACAGCCGAGCGCATGAGTGTCTGCTTCTCTGCTGCGCTGAGCTCTTCTTTTG 240
QY 354 TGGGCGAGCTTTTGGGCGAGCCCAAGATTTTCTTACCTGATGAGAGCCCGCTGCGAG 413
DB 241 TGGGCGAGCTTTTGGGCGAGCCCAAGATTTTCTTACCTGATGAGAGCCCGCTGCGAG 300
QY 414 TGTGGATGACCTTCAACAGAGACCGCGCTGATGCTGACATGAGTGTGCGGATCTGA 473
DB 301 TGTGGATGACCTTCAACAGAGACCGCGCTGATGCTGACATGAGTGTGCGGATCTGA 360
QY 474 TAGGCGCGCTTCTTGTGCGACATGAGCTCTTGTATGCTCTTCAATGAACTGGTCCCG 533
DB 361 TAGGCGCGCTTCTTGTGCGACATGAGCTCTTGTATGCTCTTCAATGAACTGGTCCCG 420
QY 534 GAGAGAGTCCAGCTCTTCAATGAGAGACAGCGGCGCTGTGTGTCGACCTGGCT 593
DB 421 GAGAGAGTCCAGCTCTTCAATGAGAGACAGCGGCGCTGTGTGTCGACCTGGCT 480
QY 594 GTGACATCATCCCAAGATGAAATATATCCCGGCGCTCACTGACAGCTCTCTGTCAGTC 653
DB 481 GTGACATCATCCCAAGATGAAATATATCCCGGCGCTCACTGACAGCTCTCTGTCAGTC 540
QY 654 AACAGCCCTTTGAGTGTGAGAGAGGCGTCCGCTCTTACAGCCACGTGGTCTCAAG 713

DB 541 AACAGCCCTTTGAGTGTGAGAGAGGCGTCCGCTCTACAGCCACGAGTGTCTCAAG 600
QY 714 AGGTGGCGCTTCTTCAACAGCGAGTCCCTTACCGCTGCTGAAGACCCCTCCACAC 773
DB 601 AGGTGGCGCTTCTTCAACAGCGAGTCCCTTACCGCTGCTGAAGACCCCTCCACAC 660
QY 774 TGCATATGCTGACCTGTCGCGGAGACCCCGGCGGTGTCCGTTCCGAGAACGACAA 833
DB 661 TGCATATGCTGACCTGTCGCGGAGACCCCGGCGGTGTCCGTTCCGAGAACGACAA 720
QY 834 AGGAGATCTCATGATTTGACAGTCCGATTTGTATGAGGCGAGCATGACGAAACTCAAG 893
DB 721 AGGAGATCTCATGATTTGACAAATCGATTGAGGCGGCAACATGAGCAAAAACACAGA 780
QY 894 AGGAGACCAACCCCTACTATGATGATGACAGTCA--TCTCCAAAGCCAGC-TGAGATCT 950
DB 781 ANAGAGACCAACCCCTACTATGATGATGAGGCGCAATCTGSCCAAGCAGCTTGAGACT 840
QY 951 ACAAGAC 958
DB 841 ACAAGAC 848

RESULT 6
AAK93921
ID AAK93921 standard; cDNA, 877 BP.
XX
AC AAK93921;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human cDNA clone representative sequence, SEQ ID NO: 2381.
DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
KW Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
QY 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesized by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX WPI: 2002-075226/10.
DR P-PSDB; AAU11274.
PT New enzyme, useful for modifying acceptor molecule, comprises an
PT isolated L-selectin sulfoltransferase-2 that directs expression of
PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
PT intestinal GlcNAc 6-sulfoltransferase
PS Claim 19, Fig 4, 98pp, English.
XX
XX The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated
CC betaal,3-N-acetylglucosaminyltransferase (betaal,3gnt) or an active
CC fragment, where betaal,3gnt directs expression of a MECA-79 antigen. The
CC invention also provides a method of treating or preventing an
CC L-selectin-mediated condition by reducing the expression or activity of a
CC betaal,3gnt that directs expression of a MECA-79 antigen. This can be done
CC by administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds betaal,3gnt,
CC and/or a betaal,3gnt antisense nucleic acid molecule. L-selectin
CC sulfoltransferase-2 (LST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of betaal,3gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents cDNA encoding human LST-2.
XX
SQ Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;
Query Match 62.1%; Score 1261.6; DB 24; Length 1333;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 155 TCTTGCAAGGCTTCCACTTACGACAAATGCTACTGCTTAAAAAATGAAGCTCTCTG 214
DB 66 TCCCTAAAGGCTTCCACTTACGACAAATGCTACTGCTTAAAAAATGAAGCTCTCTG 125
QY 215 TTCTGGTTTCCAGATGGCCATCTTGCTATTCTTCCAGATGACAGCAACAATC 274
DB 126 TTCTGGTTTCCAGATGGCCATCTTGCTATTCTTCCAGATGACAGCAACAATC 185
QY 275 AGCTCCCTGTATGAAGGACACGCCGAGCATGACGCTGCTGCTCTCTCTG 334
DB 186 AGCTCCCTGTATGAAGGACACGCCGAGCATGACGCTGCTGCTCTCTCTG 245
QY 335 CGCTCGGCTCTCTTTTGTGGGAGAGTTTGGGAGACCCAGATGTTTCTACCTG 394
DB 246 CGCTCGGCTCTCTTTTGTGGGAGAGTTTGGGAGACCCAGATGTTTCTACCTG 305
QY 395 ATGAGGCCCTCGACAGCTGTGATGACCTTCAAGCAGAGACCCGCTGATGCTGAC 454
DB 306 ATGAGGCCCTCGACAGCTGTGATGACCTTCAAGCAGAGACCCGCTGATGCTGAC 365
QY 455 ATGGGTGGGGGAGATGATACGGGCGCTCTTGTGGGAGATGAGAGCTTTGATGCC 514
DB 366 ATGGGTGGGGGAGATGATACGGGCGCTCTTGTGGGAGATGAGAGCTTTGATGCC 425
QY 515 TACATGAGACCTGTGACCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACGCCGGGCC 574
DB 426 TACATGAGACCTGTGACCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACGCCGGGCC 485
QY 575 CTGTGTGTGACACTGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCAC 634
DB 486 CTGTGTGTGACACTGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCAC 545
QY 635 TGCAAGCTCTGTGACAGCAACACCCCTTTAGGTGTGAGAGAGCTGCGCTCTCTAC 694
DB 546 TGCAAGCTCTGTGACAGCAACACCCCTTTAGGTGTGAGAGAGCTGCGCTCTCTAC 605

QY 695 AGCCACGTGTGCTCAGGAGGAGGCGCTTCTCAACCTGACATCCCTGACCCGCTGCG 754
DB 606 AGCCACGTGTGCTCAGGAGGAGGCGCTTCTTCAACCTGACATCCCTGACCCGCTGCG 665
QY 755 AAGAGCCCTCCCTCAACCTGACATGAGTGTGACCTGCTGCGGAGCCCGGCGCTGTC 814
DB 666 AAGAGCCCTCCCTCAACCTGACATGAGTGTGACCTGCTGCGGAGCCCGGCGCTGTC 725
QY 815 CGTTCGCCGAGACGCAACAGGAGATCTCATGATGTGACATGCGATGTGTGGGCGAG 874
DB 726 CGTTCGCCGAGACGCAACAGGAGATCTCATGATGTGACATGCGATGTGTGGGCGAG 785
QY 875 CATGACAGAACTCAAGAGAGAGACCAACCTACTATGTATGATGATGATCTGCCAA 934
DB 786 CATGACAGAACTCAAGAGAGAGACCAACCTACTATGTATGATGATGATGATGATGAT 845
QY 935 AGCCACTGAGATCTACAGACATCCAGTCTTGTCCCAAGGCTGTGAGAGAGCTAC 994
DB 846 AGCCACTGAGATCTACAGACATCCAGTCTTGTCCCAAGGCTGTGAGAGAGCTAC 905
QY 995 CTGCTGTGCGCTATGAGACCTGCTCGAGCCCTGTGCGCCAGACTTCCGAAATGTAT 1054
DB 906 CTGCTGTGCGCTATGAGACCTGCTCGAGCCCTGTGCGCCAGACTTCCGAAATGTAT 965
QY 1055 GAATGTGAGATGGAATTTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 1114
DB 966 GAATGTGAGATGGAATTTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 1025
QY 1115 GGCAAGGAGATGAGTACACACGCTTCCACACAAATGCGAGGATGCTTATGTCTCC 1174
DB 1026 GGCAAGGAGATGAGTACACACGCTTCCACACAAATGCGAGGATGCTTATGTCTCC 1085
QY 1175 CAGGCTTGGGCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 1234
DB 1086 CAGGCTTGGGCTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 1145
QY 1235 GATGCCATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1294
DB 1146 GATGCCATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1205
QY 1295 TTGCTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1354
DB 1206 TTGCTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1265
QY 1355 GCGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1414
DB 1266 GCGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1325
QY 1415 CTACATCT 1422
DB 1326 CTACATCT 1333
RESULT 5
AAK91803
ID AAK91803 standard; cDNA: 877 BP.
XX
XX AAK91803;
DE 06-NOV-2001 (first entry)
XX Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX Homo sapiens.
XX
XX
XX EP1130094-A2.
XX
XX
XX 05-SEP-2001.
XX
XX
XX 07-JUL-2000; 2000EP-0114089.

| | | | |
|----|------|---|------|
| QY | 414 | TGTGATACCTTTAAACAGACAGACCCCTGGATGTCGCAACAGCTGTGCGGATCTGA | 473 |
| Db | 301 | TGTGATACCTTTAAACAGACAGACCGCCTGGATGTCGCAACAGCTGTGCGGATCTGA | 360 |
| QY | 474 | TACGGGCGCTTCTTGTGCGACATGAGGCTCTTGTATGCTTACATGAAACCTGGTCCC | 533 |
| Db | 361 | TACGGGCGCTTCTTGTGCGACATGAGGCTCTTGTATGCTTACATGAAACCTGGTCCC | 420 |
| QY | 534 | GGAGACATGCCAGCTTTTCTAGAGGGAGAAACCGCGGGCCGTGTCTGTGCAACCTGCT | 593 |
| Db | 421 | GGAGACATGCCAGCTTTTCTAGAGGGAGAAACCGCGGGCCGTGTCTGTGCAACCTGCT | 480 |
| QY | 554 | GTGACATCATCCCAAGATGAATCATCCCGCGGCTCACTGCACAGGCTCTGTGCAGTC | 653 |
| Db | 481 | GTGACATCATCCCAAGATGAATCATCCCGCGGCTCACTGCAGGCTCTGTGCAGTC | 540 |
| QY | 654 | AACAGCCCTTTSAGAGTGTGGAGAAAGGCGCTGCGCTCTCATAACGCACAGGAGTCAAG | 713 |
| Db | 541 | AACAGCCCTTTSAGAGTGTGGAGAAAGGCGCTGCGCTCTCATAACGCACAGGAGTCAAG | 600 |
| QY | 714 | AGGTGCGCTTTTCAACCTCAGTCCCTCAGTCCCTCAGTGGTGAAGAACCCCTCCCTCAAC | 773 |
| Db | 601 | AGGTGCGCTTTTCAACCTCAGTCCCTCAGTCCCTCAGTGGTGAAGAACCCCTCCCTCAAC | 660 |
| QY | 774 | TGCATATGCTCACCTGCTGTCGGGACCCCCCGGGCCGTGTCCGTTGCCAGAAACGCACAA | 833 |
| Db | 661 | TGCATATGCTCACCTGCTGTCGGGACCCCCCGGGCCGTGTCCGTTGCCAGAAACGCACAA | 720 |
| QY | 834 | AGGAGATATCTCATATTTAGACAGTCGATTTGTATGAGGGAGCATGAGAGCAAACTCAAGA | 893 |
| Db | 721 | AGGAGATATCTCATATTTAGACAGTCGATTTGTATGAGGGAGCATGAGAGCAAACTCAAGA | 780 |
| QY | 894 | AGGAGAGACCAACCTACTATATGTATGATGAGTCACTTGGCCAAAGCAGCTGGAGATCTTACA | 953 |
| Db | 781 | AGGAGAGACCAACCTACTATATGTATGATGAGTCACTTGGCCAAAGCAGCTGGAGATCTTACA | 840 |
| QY | 954 | AGACCATCAAGTCTTCCCTCCCAAGGCCCTGAGAGAAAGCTACAGCTTGTGCCCTATGAG | 1012 |
| Db | 841 | AGACCATCAAGTCTTCCCTCCCAAGGCCCTGAGAGAAAGCTACAGCTTGTGCCCTATGAG | 900 |
| QY | 1014 | ACTGTGCTCGAGCCCTGTGAGCCAGACTTCCCGAATGATGAAATTCGTGGATTGGAAAT | 1077 |
| Db | 901 | ACTGTGCTCGAGCCCTGTGAGCCAGACTTCCCGAATGATGAAATTCGTGGATTGGAAAT | 960 |
| QY | 1074 | TCTTGGCCCCATCTCAACACTGGTGGATPAATCATCCCGAGGCAAGGGCATGGGTGAC | 1133 |
| Db | 961 | TCTTGGCCCCATCTCAACACTGGTGGATPAATCATCCCGAGGCAAGGGCATGGGTGAC | 1020 |
| QY | 1134 | ACGCTTTCACACAAATGCCAGGAGTACCCTTATGTCTCTCCAGGCTTGGCCGCTGCTT | 1193 |
| Db | 1021 | ACGCTTTCACACAAATGCCAGGAGTACCCTTATGTCTCTCCAGGCTTGGCCGCTGCTT | 1080 |
| QY | 1194 | TGGCCCTTGAAGAAAGTTTCTCGACTTCAGAAACCGTGTGGCATGCGATGAATTGGACGG | 1255 |
| Db | 1081 | TGGCCCTTGAAGAAAGTTTCTCGACTTCAGAAACCGTGTGGCATGCGATGAATTGGACGG | 1140 |
| QY | 1254 | GCTACCGCACGTGATCTGAACAAGAACAGAAACCTGTTGCTGATCTCTTCTGTCTA | 1313 |
| Db | 1141 | GCTACCGCACGTGATCTGAACAAGAACAGAAACCTGTTGCTGATCTCTTCTGTCTA | 1200 |
| QY | 1314 | CCTGAGCATGTGCCGAGACAAATCCACTAAGAGGTTGABAAGGCTTGTGTCGACTTGGT | 1373 |
| Db | 1201 | CCTGAGCATGTGCCGAGACAAATCCACTAAGAGGTTGABAAGGCTTGTGTCGACTTGGT | 1260 |
| QY | 1374 | GTCAGCCTCAGTCACTTCTCTGATAGTCTTGTAGGCTTGCCTTACATCTCTAGAGCTTAA | 1433 |
| Db | 1261 | GTCAGCCTCAGTCACTTCTCTGATAGTCTTGTAGGCTTGCCTTACATCTCTAGAGCTTAA | 1320 |
| QY | 1434 | CTAATATCTCTGTGGATATCAACATGATGTGATTTGTCTCACAGCTGCTCAAGCAGAG | 1493 |
| Db | 1321 | CTAATATCTCTGTGGATATCAACATGATGTGATTTGTCTCACAGCTGCTCAAGCAGAG | 1380 |

| Chr | Start (Mb) | End (Mb) | Gene | Transcript | Accession | Length (bp) | GC (%) | GC3 (%) | GC4 (%) | GC5 (%) | GC6 (%) | GC7 (%) | GC8 (%) | GC9 (%) | GC10 (%) | GC11 (%) | GC12 (%) | GC13 (%) | GC14 (%) | GC15 (%) | GC16 (%) | GC17 (%) | GC18 (%) | GC19 (%) | GC20 (%) | GC21 (%) | GC22 (%) | GC23 (%) | GC24 (%) | GC25 (%) | GC26 (%) | GC27 (%) | GC28 (%) | GC29 (%) | GC30 (%) | GC31 (%) | GC32 (%) | GC33 (%) | GC34 (%) | GC35 (%) | GC36 (%) | GC37 (%) | GC38 (%) | GC39 (%) | GC40 (%) | GC41 (%) | GC42 (%) | GC43 (%) | GC44 (%) | GC45 (%) | GC46 (%) | GC47 (%) | GC48 (%) | GC49 (%) | GC50 (%) | GC51 (%) | GC52 (%) | GC53 (%) | GC54 (%) | GC55 (%) | GC56 (%) | GC57 (%) | GC58 (%) | GC59 (%) | GC60 (%) | GC61 (%) | GC62 (%) | GC63 (%) | GC64 (%) | GC65 (%) | GC66 (%) | GC67 (%) | GC68 (%) | GC69 (%) | GC70 (%) | GC71 (%) | GC72 (%) | GC73 (%) | GC74 (%) | GC75 (%) | GC76 (%) | GC77 (%) | GC78 (%) | GC79 (%) | GC80 (%) | GC81 (%) | GC82 (%) | GC83 (%) | GC84 (%) | GC85 (%) | GC86 (%) | GC87 (%) | GC88 (%) | GC89 (%) | GC90 (%) | GC91 (%) | GC92 (%) | GC93 (%) | GC94 (%) | GC95 (%) | GC96 (%) | GC97 (%) | GC98 (%) | GC99 (%) | GC100 (%) | GC101 (%) | GC102 (%) | GC103 (%) | GC104 (%) | GC105 (%) | GC106 (%) | GC107 (%) | GC108 (%) | GC109 (%) | GC110 (%) | GC111 (%) | GC112 (%) | GC113 (%) | GC114 (%) | GC115 (%) | GC116 (%) | GC117 (%) | GC118 (%) | GC119 (%) | GC120 (%) | GC121 (%) | GC122 (%) | GC123 (%) | GC124 (%) | GC125 (%) | GC126 (%) | GC127 (%) | GC128 (%) | GC129 (%) | GC130 (%) | GC131 (%) | GC132 (%) | GC133 (%) | GC134 (%) | GC135 (%) | GC136 (%) | GC137 (%) | GC138 (%) | GC139 (%) | GC140 (%) | GC141 (%) | GC142 (%) | GC143 (%) | GC144 (%) | GC145 (%) | GC146 (%) | GC147 (%) | GC148 (%) | GC149 (%) | GC150 (%) | GC151 (%) | GC152 (%) | GC153 (%) | GC154 (%) | GC155 (%) | GC156 (%) | GC157 (%) | GC158 (%) | GC159 (%) | GC160 (%) | GC161 (%) | GC162 (%) | GC163 (%) | GC164 (%) | GC165 (%) | GC166 (%) | GC167 (%) | GC168 (%) | GC169 (%) | GC170 (%) | GC171 (%) | GC172 (%) | GC173 (%) | GC174 (%) | GC175 (%) | GC176 (%) | GC177 (%) | GC178 (%) | GC179 (%) | GC180 (%) | GC181 (%) | GC182 (%) | GC183 (%) | GC184 (%) | GC185 (%) | GC186 (%) | GC187 (%) | GC188 (%) | GC189 (%) | GC190 (%) | GC191 (%) | GC192 (%) | GC193 (%) | GC194 (%) | GC195 (%) | GC196 (%) | GC197 (%) | GC198 (%) | GC199 (%) | GC200 (%) | GC201 (%) | GC202 (%) | GC203 (%) | GC204 (%) | GC205 (%) | GC206 (%) | GC207 (%) | GC208 (%) | GC209 (%) | GC210 (%) | GC211 (%) | GC212 (%) | GC213 (%) | GC214 (%) | GC215 (%) | GC216 (%) | GC217 (%) | GC218 (%) | GC219 (%) | GC220 (%) | GC221 (%) | GC222 (%) | GC223 (%) | GC224 (%) | GC225 (%) | GC226 (%) | GC227 (%) | GC228 (%) | GC229 (%) | GC230 (%) | GC231 (%) | GC232 (%) | GC233 (%) | GC234 (%) | GC235 (%) | GC236 (%) | GC237 (%) | GC238 (%) | GC239 (%) | GC240 (%) | GC241 (%) | GC242 (%) | GC243 (%) | GC244 (%) | GC245 (%) | GC246 (%) | GC247 (%) | GC248 (%) | GC249 (%) | GC250 (%) | GC251 (%) | GC252 (%) | GC253 (%) | GC254 (%) | GC255 (%) | GC256 (%) | GC257 (%) | GC258 (%) | GC259 (%) | GC260 (%) | GC261 (%) | GC262 (%) | GC263 (%) | GC264 (%) | GC265 (%) | GC266 (%) | GC267 (%) | GC268 (%) | GC269 (%) | GC270 (%) | GC271 (%) | GC272 (%) | GC273 (%) | GC274 (%) | GC275 (%) | GC276 (%) | GC277 (%) | GC278 (%) | GC279 (%) | GC280 (%) | GC281 (%) | GC282 (%) | GC283 (%) | GC284 (%) | GC285 (%) | GC286 (%) | GC287 (%) | GC288 (%) | GC289 (%) | GC290 (%) | GC291 (%) | GC292 (%) | GC293 (%) | GC294 (%) | GC295 (%) | GC296 (%) | GC297 (%) | GC298 (%) | GC299 (%) | GC300 (%) | GC301 (%) | GC302 (%) | GC303 (%) | GC304 (%) | GC305 (%) | GC306 (%) | GC307 (%) | GC308 (%) | GC309 (%) | GC310 (%) | GC311 (%) | GC312 (%) | GC313 (%) | GC314 (%) | GC315 (%) | GC316 (%) | GC317 (%) | GC318 (%) | GC319 (%) | GC320 (%) | GC321 (%) | GC322 (%) | GC323 (%) | GC324 (%) | GC325 (%) | GC326 (%) | GC327 (%) | GC328 (%) | GC329 (%) | GC330 (%) | GC331 (%) | GC332 (%) | GC333 (%) | GC334 (%) | GC335 (%) | GC336 (%) | GC337 (%) | GC338 (%) | GC339 (%) | GC340 (%) | GC341 (%) | GC342 (%) | GC343 (%) | GC344 (%) | GC345 (%) | GC346 (%) | GC347 (%) | GC348 (%) | GC349 (%) | GC350 (%) | GC351 (%) | GC352 (%) | GC353 (%) | GC354 (%) | GC355 (%) | GC356 (%) | GC357 (%) | GC358 (%) | GC359 (%) | GC360 (%) | GC361 (%) | GC362 (%) | GC363 (%) | GC364 (%) | GC365 (%) | GC366 (%) | GC367 (%) | GC368 (%) | GC369 (%) | GC370 (%) | GC371 (%) | GC372 (%) | GC373 (%) | GC374 (%) | GC375 (%) | GC376 (%) | GC377 (%) | GC378 (%) | GC379 (%) | GC380 (%) | GC381 (%) | GC382 (%) | GC383 (%) | GC384 (%) | GC385 (%) | GC386 (%) | GC387 (%) | GC388 (%) | GC389 (%) | GC390 (%) | GC391 (%) | GC392 (%) | GC393 (%) | GC394 (%) | GC395 (%) | GC396 (%) | GC397 (%) | GC398 (%) | GC399 (%) | GC400 (%) | GC401 (%) | GC402 (%) | GC403 (%) | GC404 (%) | GC405 (%) | GC406 (%) | GC407 (%) | GC408 (%) | GC409 (%) | GC410 (%) | GC411 (%) | GC412 (%) | GC413 (%) | GC414 (%) | GC415 (%) |
|-----|------------|----------|------|------------|-----------|-------------|--------|---------|---------|---------|---------|---------|---------|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|-----|------------|----------|------|------------|-----------|-------------|--------|---------|---------|---------|---------|---------|---------|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

| | | | |
|----|------|---|------|
| Db | 960 | TTGGCCCAAGCCCTCGAGGAAAGCCTACCTGGCTTGGCGCTATGAGGACCTGGACTGAGACC | 1015 |
| QY | 1028 | CTGTGGCCCCAGACCTTCCCGAATGATATGAATTCGTGGGATTTGGAAATTTCTTGGCCCATCTT | 1087 |
| Db | 1020 | CTGTGGCCCCAGACTTCCCGAATGATGAAATTCGTGGGATTTGGAAATTTCTTGGCCCATCTT | 1079 |
| QY | 1088 | CAGACCTGGGGTGAATACATCAATCCGACGACGAGGCGATGGGTGACACCGCTTCCACACA | 1147 |
| Db | 1080 | CAGACCTGGGGTGAATACATCAATCCGACGACGAGGCGATGGGTGACACCGCTTCCACACA | 1139 |
| QY | 1148 | AATGCCAGGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGTCTTTGGCCCTATGAAAAG | 1207 |
| Db | 1140 | AATGCCAGGGATGCCCTTAATGTCTCCAGGCTTGGCGCTGTCTTTGGCCCTATGAAAAG | 1199 |
| QY | 1208 | GTTCCTCGAATTCAGAAAGCCTGGGCGATGCGCATGAAATTTGCTGGGCTACCGGCACGTC | 1267 |
| Db | 1200 | GTTCCTCGAATTCAGAAAGCCTGGGCGATGCGCATGAAATTTGCTGGGCTACCGGCACGTC | 1259 |
| QY | 1268 | AGATCTGGAACAGAACAGAGAAACCTGGTGGTGGATCTTCTGTCTACCTGGAAGTGTCCCT | 1327 |
| Db | 1260 | AGATCTGAACAGAACAGAGAAACCTGGTGGTGGATCTTCTGTCTACCTGGAAGTGTCCCT | 1319 |
| QY | 1328 | GAGCAAAATCCACTTAAGAGGTTTAGAAAGCTTTGCTGCCACTGGGTGTCAGGCTTCAGTCA | 1387 |
| Db | 1320 | GAGCAAAATCCACTTAAGAGGTTTAGAAAGCTTTGCTGCCACTGGGTGTCAGGCTTCAGTCA | 1379 |
| QY | 1388 | CTTTCCTGTGATCTCTCTGAGCCCTTGACCTCCATACATCTGTGAGCCTTAACATGATGTGTGG | 1447 |
| Db | 1380 | CTTTCCTGTGATCTCTCTGAGCCCTTGACCTCCATACATCTGTGAGCCTTAACATGATGTGTGG | 1439 |
| QY | 1448 | GTATCACACTGAGTGTGATGTTGTGTCCACAGCTGCTCAAGCAGAAAGACTTTTGTGTCCA | 1507 |
| Db | 1440 | GTATCACACTGAGTGTGATGTTGTGTCCACAGCTGCTCAAGCAGAAAGACTTTTGTGTCCA | 1499 |
| QY | 1508 | TGCTGTGTGTAGAAAACACACTGGGGAACTATGTAGCAGCAACATCCGACACGTGTAA | 1567 |
| Db | 1500 | TGCTGTGTGTAGAAAACACACTGGGGAACTATGTAGCAGCAACATCCGACACGTGTAA | 1559 |
| QY | 1568 | ACAGGATATGCTCTCTCTTCTTTCTTGATCTTCTGTGATCTGTGTGGGACAGCTTCAGACATTT | 1627 |
| Db | 1560 | ACAGGATATGCTCTCTCTTCTTTCTTGATCTTCTGTGATCTGTGTGGGACAGCTTCAGACATTT | 1619 |
| QY | 1628 | GTGGCCCTGGAGGCGCTATTAGCACAGCAACAGATATCGTGGAAATTAATCCATTAACCTGCC | 1687 |
| Db | 1620 | GTGGCCCTGGAGGCGCTATTAGCACAGCAACAGATATCGTGGAAATTAATCCATTAACCTGCC | 1679 |
| QY | 1688 | TGTCACCATCTTGGCCCATGGGGAATGGATCTTTACCAAAAGAGCTCACACACATTTTCC | 1747 |
| Db | 1680 | TGTCACCATCTTGGCCCATGGGGAATGGATCTTTACCAAAAGAGCTCACACACATTTTCC | 1739 |
| QY | 1748 | ACAAAGATGCAAAATTCAGACCCCTTGGAGTTTCCCAATGSAATTCAGAGGAAGATGGGAA | 1807 |
| Db | 1740 | ACAAAGATGCAAAATTCAGACCCCTTGGAGTTTCCCAATGSAATTCAGAGGAAGATGGGAA | 1799 |
| QY | 1808 | CAGAGTGGATGCTTACTATAGACTTGACATCAACATCAAGCTATCGGTAATCAGAAATATAGA | 1867 |
| Db | 1800 | CAGAGTGGATGCTTACTATAGACTTGACATCAACATCAAGCTATCGGTAATCAGAAATATAGA | 1855 |
| QY | 1868 | AACAAAAATCTCTGCAAAAAGAGCAAGCTTTAAGTTCACAGAGGTGCTGGGCTGCATTT | 1927 |
| Db | 1860 | AACAAAAATCTCTGCAAAAAGAGCAAGCTTTAAGTTCACAGAGGTGCTGGGCTGCATTT | 1919 |
| QY | 1928 | GAAATATCACTCCCTCCCTGATTTTCCATCAATCAATGAGAGACTTTGACCTGTGAAGCTGC | 1987 |
| Db | 1920 | GAAATATCACTCCCTCCCTGATTTTCCATCAATCAATGAGAGACTTTGACCTGTGAAGCTGC | 1979 |
| QY | 1988 | CATCTGTTAATACTAATAATCCCAAAATAG | 2017 |
| Db | 1980 | CATCTGTTAATACTAATAATCCCAAAATAG | 2009 |

| | | | | |
|----|---|--|---------------|----------|
| ID | AAK94229 | standard: | cdna: | 1979 BP. |
| XX | AC | AAK94229; | | |
| XX | DT | 06-NOV-2001 | (first entry) | |
| DE | Human full-length cDNA, SEQ ID NO: 2816. | | | |
| XX | Human; full length CDNA; cDNA synthesis; oligo-capping; ss. | | | |
| OS | Homo sapiens. | | | |
| XX | EP1130094-A2. | | | |
| PN | 05-SEP-2001. | | | |
| PD | 07-JUL-2000; 2000EP-0114089. | | | |
| PF | 08-JUL-1999; 99JP-0194486. | | | |
| XX | PR 11-JAN-2000; 2000JP-0118774. | | | |
| XX | PR 02-MAY-2000; 2000JP-0183765. | | | |
| XX | (HELI-) HELIX RES INST. | | | |
| PA | Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y; | | | |
| XX | Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H; | | | |
| XX | WPI: 2001-524255/58. | | | |
| DR | P-Psde; AAM93309. | | | |
| FT | 830 Primers useful for synthesizing full length cDNA clones and their | | | |
| PT | use in genetic manipulation - | | | |
| PS | Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English. | | | |
| CC | The invention relates to primers for synthesizing full length cDNA | | | |
| CC | clones. 830 cDNA molecules encoding a human protein have been | | | |
| CC | isolated and nucleotide sequences of 5' and 3' ends of the cDNA | | | |
| CC | molecules have been determined. Primers for synthesizing the full length | | | |
| CC | cDNA are useful for clarifying the function of the protein encoded by | | | |
| CC | the cDNA. The full length clones were obtained by construction of full | | | |
| CC | length enriched cDNA libraries that were synthesized by the oligo-capping | | | |
| CC | method. The primers enable the production of the full length cDNA easily | | | |
| CC | without any special methods. The present sequence is a full length | | | |
| CC | human cDNA of the invention. | | | |
| CC | Note: The sequence data for this patent did not form part of the printed | | | |
| CC | specification, but was obtained in CD-ROM format directly from EPO. | | | |
| XX | Sequence 1979 BP: 451 A; 542 C; 474 G; 512 T; 0 other: | | | |
| SQ | | | | |
| | Query Match | 93.5%; Score 1900.8; DB 22; Length 1979; | | |
| | Best Local Similarity | 99.9%; Pred. No. 0; | | |
| | Matches 1902; Conservative | 0; Mismatches | 2; Indels | 0; Gaps |
| | | | | 0. |
| QY | 114 ACTTCACAGCTCCTCGGAGGAGGTGCTTCTCAAGCCGGTTGCAGAGCTTCCACT | 173 | | |
| DY | 1 ACTTTCACAGCTCCTCGGAGGAGGTGCTTCTCAAGCCGGTTGCAGAGCTTCCACT | 60 | | |
| DY | 174 TCAGCACAAATGCTACTGCTATAAAAAAATGAAGCTCCTGCTGTTTGCTTCCCAGATGG | 233 | | |
| DY | 61 TCAGCACAAATGCTACTGCTATAAAAAAATGAAGCTCCTGCTGTTTGCTTCCCAGATGG | 120 | | |
| DY | 234 CCATCTGGGCTATTCTTCACAAATGTAAGGACACAACATCAGTCCCTGTATGAAG | 293 | | |
| DY | 121 CCATCTGGGCTATTCTTCACAAATGTAAGGACACAACATCAGTCCCTGTATGAAG | 180 | | |
| DY | 294 CACAGCCGAGCGCATGACAGTGTGTTCTGCTTCCGCGGCTCTGCTCTTTTG | 353 | | |
| DY | 181 CACAGCCGAGCGCATGACAGTGTGTTCTGCTTCCGCGGCTCTGCTCTTTTG | 240 | | |
| DY | 354 TGCGGACAGTTTTTGGGACGACACCAGATGTTTTTACTACTATGAGCCCCGCTGGCAG | 413 | | |
| DY | 241 TGGGGACAGTTTTTGGGACGACACCAGATGTTTTTACTACTATGAGCCCCGCTGGCAG | 300 | | |

Db 1861 AATATGAAACAAATCTCTGCACAAAGAGCAAGCTTTAAGTTCACAGGGTGGCTGGGC 1920
 QY 1921 TGCATTTGAATATCACTTCCCTCTGCATTTTCCCATCATAGACACTTGAACCTGTG 1980
 Db 1921 TGCATTTGAATATCACTTCCCTCTGCATTTTCCCATCATAGACACTTGAACCTGTG 1980
 QY 1981 AAGCTGCCATCTGTATTAATCTAAATTTCCCAATATAGAAAAA 2032
 Db 1981 AAGCTGCCATCTGTATTAATCTAAATTTCCCAATATAGAAAAA 2032
 RESULT 2
 AA294211
 ID AA294211 standard; cDNA: 2065 BP.
 AC AA294211;
 DT 19-JUN-2000 (first entry)
 XX Human transferase TRNSFS-11 cDNA clone 2617407CBL.
 XX
 XX Transferase; TRNSFS-11; human; antitumor; cell proliferation;
 XX inflammation; gastrointestinal disorder; developmental disorder;
 XX genetic disorder; neurological disorder; reproductive disorder;
 XX smooth muscle disorder; immunological disorder; gene therapy;
 XX diagnosis; N-acetylglucosamine 6-O-sulfotransferase; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 174..1334
 FT /*tag= a
 PN WO200014251-A2.
 XX 16-MAR-2000.
 XX 09-SEP-1999; 99WO-US20989.
 XX 10-SEP-1998; 98US-0150657.
 XX 04-NOV-1998; 98US-0186779.
 XX 11-MAY-1999; 99US-0133642.
 PA (INCYTE PHARM INC.
 PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
 PI Hillman JL, Azimzai Y;
 DR WPI: 2000-256996/22.
 DR P-PDB: AAY79219.
 XX Human transferase proteins useful for preventing, diagnosing and
 XX treating cancers and developmental, gastrointestinal, genetic,
 XX immunological, neurological, reproductive and smooth muscle disorders -
 PS Claim 9; Page 104-105; 113pp; English.
 XX The present sequence is that of cDNA clone 2617407CBL encoding
 XX human transferase TRNSFS-11 (see AAY79219), 1 of 15 cloned human
 XX transferase proteins of the invention (see AAY79209-23). The clone
 XX is isolated from gall bladder cDNA library GBLANOT01. TRNSFS-11
 XX is expressed in dermatologic and gastrointestinal tissues,
 XX especially those associated with inflammation and cell
 XX proliferation. It shows homology to mouse N-acetylglucosamine
 XX 6-O-sulfotransferase. A polynucleotide comprising nucleotides
 XX 264-333 or 1272-1331 of the present sequence can be used as a DNA
 XX probe. The new human transferases and polynucleotides can be used
 XX in the diagnosis, prevention and treatment (including gene therapy
 XX and antisense therapy) of cancer, developmental disorders,
 XX gastrointestinal disorders, genetic disorders, immunological
 XX disorders, neurological disorders, reproductive disorders, and
 XX smooth muscle disorders.

XX SQ Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other;
 Query Match 97.5%; Score 1981.7; DB 21; Length 2065;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2005; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
 QY 9 GCCAGATGCTCCCTCAGTCTGCGGGAATGCTTCTCATTGGCTGCTCCAGCCACCTC 68
 Db 1 GCCAGATGCTCCCTCAGTCTGCGGGAATGCTTCTCATTGGCTGCTCCAGCCACCTC 60
 QY 69 AAGCAGTCTCCCTCCAGCCCTTGATGCTCAGCAGTGAAGCTTACTTTCACAGCTTCC 128
 Db 61 AAGCAGTCTCCCTCCAGCCCTTGATGCTCAGCAGTGAAGCTTACTTTCACAGCTTCC 120
 QY 129 TGGAGAGAGTGTCTTCTCAAGCCCTCTGCAAGGCTTCCACTTCAGACAAATGCTAC 188
 Db 121 TGGAGAGAGTGTCTTCTCAAGCCCTCTGCAAGGCTTCCACTTCAGACAAATGCTAC 180
 QY 189 TGGCTAAAAAATGAAAGCTCTCTCTTCTGTTTCCAGATGGCATTGGCTCTAT 248
 Db 181 TGGCTAAAAAATGAAAGCTCTCTCTTCTGTTTCCAGATGGCATTGGCTCTAT 240
 QY 249 TCTTCCACATGTACACGCCAACATCAGCTCCCTGTCTATGAAGGACAGCCGAGCGCA 308
 Db 241 TCTTCCACATGTACACGCCAACATCAGCTCCCTGTCTATGAAGGACAGCCGAGCGCA 300
 QY 309 TGCACGTGCTGTTCTGTCTCTCTCTGCGCTCTGCGCTCTTCTTTTGGGGCAGCTTTTG 368
 Db 301 TGCACGTGCTGTTCTGTCTCTCTCTGCGCTCTGCGCTCTTCTTTTGGGGCAGCTTTTG 360
 QY 369 GGCAGACCCAGATGTTTCTTCTACCTGATGAGGCGCCCTGGGACGATGATGACCTTCA 428
 Db 361 GGCAGACCCAGATGTTTCTTCTACCTGATGAGGCGCCCTGGGACGATGATGACCTTCA 420
 QY 429 AGCAGAGCACCGCTGATGCTGACATGCTGTGCGGATGATACGGGCGCTTCT 488
 Db 421 AGCAGAGCACCGCTGATGCTGACATGCTGTGCGGATGATACGGGCGCTTCT 480
 QY 489 TGTGCGACATGAGCGCTTGTGATGCTTACATGAACTGCTCCCGGAGACAGTCCAGCC 548
 Db 481 TGTGCGACATGAGCGCTTGTGATGCTTACATGAACTGCTCCCGGAGACAGTCCAGCC 540
 QY 549 TCTTTCAGTGGGGAACACAGCCGCGCTGTCTGTCACCTCCCTGTCATCATCATCCAC 608
 Db 541 TCTTTCAGTGGGGAACACAGCCGCGCTGTCTGTCACCTCCCTGTCATCATCATCCAC 600
 QY 609 AAGATGAATATCCTCCCGCGCTCAGTCAAGCTCTGTGAGTCAACAGCCCTTTC-AG 667
 Db 601 AAGATG-AATTCATCCCGCGCTCAGTCAAGCTCTGTGAGTCAACAGCCCTTTCAGG 659
 QY 668 GTGTGTGGAAGAAGGCTGCGGCTGCTCTACAGCCAGCGTGTGAGAGGTGGCTTCTTC 727
 Db 660 TTTTGTGGAAGAAGGCTGCGGCTGCTCTACAGCCAGCGTGTGAGAGGTGGCTTCTTC 719
 QY 728 AACCTGCACTCCCTTACCCGCTGCTGAAGAACCCTCCCTCAACCTGCATATGCTGCAC 787
 Db 720 AACCTGCACTCCCTTACCCGCTGCTGAAGAACCCTCCCTCAACCTGCATATGCTGCAC 779
 QY 788 CTGTGTCGGGAGACCCCGCGGCTGTCTGCTCCGAGAAAGCAGCAAGGAGATCTCATG 847
 Db 780 CTGTGTCGGGAGACCCCGCGGCTGTCTGCTCCGAGAAAGCAGCAAGGAGATCTCATG 839
 QY 848 ATTGACAGTGCATTTGTGATGGGAGCATGAGCAGAAATCTCAAGAGAGAGACCAACCC 907
 Db 840 ATTGACAGTGCATTTGTGATGGGAGCATGAGCAGAAATCTCAAGAGAGAGACCAACCC 899
 QY 908 TACTATGTATGCAAGTCTATCTGCCAAGCAGCTGGAGATTTAACAAGCATTCAGTCC 967
 Db 900 TACTATGTATGCAAGTCTATCTGCCAAGCAGCTGGAGATTTAACAAGCATTCAGTCC 959
 QY 968 TTGCCAAGGCGCTGAGAGACGTACCTGCTTGTGCGTATGAGACCTGGCTGCAGCC 1027

XX Human and murine glycosyl sulfotransferase 3 and related
PT polynucleotides
XX
PS Claim 4; Fig 1; 59pp; English.
XX
CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.
XX
SQ Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;
Query Match 100.0%; Score 2032; DB 20; Length 2032;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCTGAGGCCAGGATCCCTCCAGCTGGGGGAAATGCTTCTCATTTGCTTCCAG 60
DB 1 GGCCTGAGGCCAGGATCCCTCCAGCTGGGGGAAATGCTTCTCATTTGCTTCCAG 60
QY 61 CCCAGCTCAAGAGAGTCCCGACCCCTTGAAGTCCAGAGCTTAAAGCTTACTTCA 120
DB 61 CCCAGCTCAAGAGAGTCCCGACCCCTTGAAGTCCAGAGCTTAAAGCTTACTTCA 120
QY 121 CAGCTTCTGGAGAGCGAGTCTTCTCAAGCCGCTTTCAGAGTCTTCCACTTCAGCAC 180
DB 121 CAGCTTCTGGAGAGCGAGTCTTCTCAAGCCGCTTTCAGAGTCTTCCACTTCAGCAC 180
QY 181 AATCTCTACTGCTTAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCCATCTT 240
DB 181 AATCTCTACTGCTTAAATGAAGCTCTGCTGTTTCTGTTTCCAGATGGCCATCTT 240
QY 241 GGCCTCTATCTTCCACATGTACAGCCACAAATCAGCTCCCTGCTATGAGAGCACGCC 300
DB 241 GGCCTCTATCTTCCACATGTACAGCCACAAATCAGCTCCCTGCTATGAGAGCACGCC 300
QY 301 CGAGCGATGACGCTGCTGTTCTTCTTCTTCTGCGCTCTGCGCTCTTTTGTGGGCA 360
DB 301 CGAGCGATGACGCTGCTGTTCTTCTTCTTCTTCTGCGCTCTGCGCTCTTTTGTGGGCA 360
QY 361 GCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAGAGCCGCTGGACAGTGTGAT 420
DB 361 GCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAGAGCCGCTGGACAGTGTGAT 420
QY 421 GACCTTCAGCAGAGCAACCGCTGATGTGACATGCTGCTGGGAGTGTATACGGGC 480
DB 421 GACCTTCAGCAGAGCAACCGCTGATGTGACATGCTGCTGGGAGTGTATACGGGC 480
QY 481 CGTCTCTGTGGAGACATGAGCGCTTGTGATGCTTACATGAGAACTGTCCCGAGACA 540
DB 481 CGTCTCTGTGGAGACATGAGCGCTTGTGATGCTTACATGAGAACTGTCCCGAGACA 540
QY 541 GTCAGAGCTCTTTCAGTGGAGAACAGCCGGGCGCTGTGTTCTGACACTGCTGTGACAT 600
DB 541 GTCAGAGCTCTTTCAGTGGAGAACAGCCGGGCGCTGTGTTCTGACACTGCTGTGACAT 600
QY 601 CATCCCAAAATGAATCATCCCGCGGCTCACTGAGGCTCTGTGCACTGCTGTGACAT 660
DB 601 CATCCCAAAATGAATCATCCCGCGGCTCACTGAGGCTCTGTGCACTGCTGTGACAT 660
QY 661 CTTGAGGTGGTGAAGAGGCTCCGCTCTTACAGCCAGTGTGCTCAAGAGGTGGC 720
DB 661 CTTGAGGTGGTGAAGAGGCTCCGCTCTTACAGCCAGTGTGCTCAAGAGGTGGC 720
QY 721 CTTCTTCACTGCAAGTCCCTTACCGCGTGTCTAAGAGCCCTTCCCTCAACCTGATAT 780
DB 721 CTTCTTCACTGCAAGTCCCTTACCGCGTGTCTAAGAGCCCTTCCCTCAACCTGATAT 780

QY 781 CGTGCACCTGTGTCCGGGAGCCCCCGGCGGTGTTCCGTTCCGAGAACCCACAAGGGAGA 840
DB 781 CGTGCACCTGTGTCCGGGAGCCCCCGGCGGTGTTCCGTTCCGAGAACCCACAAGGGAGA 840
QY 841 TCTCATGATGACAGTCCGATGTGATGGGGAGCATGAGTGAAGAACTCAAGAAAGAGAGA 900
DB 841 TCTCATGATGACAGTCCGATGTGATGGGGAGCATGAGTGAAGAACTCAAGAAAGAGAGA 900
QY 901 CCAACCCCTACTATGTGATGACAGTCAATCGCCAAAGCCAGCTGAGATCTACACACCAT 960
DB 901 CCAACCCCTACTATGTGATGACAGTCAATCGCCAAAGCCAGCTGAGATCTACACACCAT 960
QY 961 CCAGTCTTGCCCAAGGCGCTCAGAAAGCATCTGCTGCGCTATGAGAGCTTGGC 1020
DB 961 CCAGTCTTGCCCAAGGCGCTCAGAAAGCATCTGCTGCGCTATGAGAGCTTGGC 1020
QY 1021 TCGAGCCCTGTGGCCAGACTTCCGAAATGATGATATGATGAGATGGAATTCCTTGC 1080
DB 1021 TCGAGCCCTGTGGCCAGACTTCCGAAATGATGATATGATGAGATGGAATTCCTTGC 1080
QY 1081 CCATCTTCAGACTGGGTGCATATCATCACCCGAGGCAAGGSCATGGGTGACAGCTTT 1140
DB 1081 CCATCTTCAGACTGGGTGCATATCATCACCCGAGGCAAGGSCATGGGTGACAGCTTT 1140
QY 1141 CCACACAATGCGAGGAGTGCCTTAATGTCTCCAGAGGCTTGGCTGTGCTTCCCTA 1200
DB 1141 CCACACAATGCGAGGAGTGCCTTAATGTCTCCAGAGGCTTGGCTGTGCTTCCCTA 1200
QY 1201 TGAAGAGTTTCTCGACTTCAGAAAGCTGTGGGATGCGATGAATTTGCTGGCTACCG 1260
DB 1201 TGAAGAGTTTCTCGACTTCAGAAAGCTGTGGGATGCGATGAATTTGCTGGCTACCG 1260
QY 1261 CCACGTCATATCGAACAAGACAGAAACCCGTTGCTGAGTCTTCTGCTTACCTGGAC 1320
DB 1261 CCACGTCATATCGAACAAGACAGAAACCCGTTGCTGAGTCTTCTGCTTACCTGGAC 1320
QY 1321 TGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTTGTGTCACACTGCTGTGAGCC 1380
DB 1321 TGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTTGTGTCACACTGCTGTGAGCC 1380
QY 1381 TCAGTACTTCTCTAATGCTTCTGAGCCTTCCCTACATCTCTGAGCCTTAACATCAG 1440
DB 1381 TCAGTACTTCTCTAATGCTTCTGAGCCTTCCCTACATCTCTGAGCCTTAACATCAG 1440
QY 1441 TCTGTGGATATCACTAGTGTGAGTGTGTGTCACACCTGCTCAAGCAGAGGACTTTT 1500
DB 1441 TCTGTGGATATCACTAGTGTGAGTGTGTGTCACACCTGCTCAAGCAGAGGACTTTT 1500
QY 1501 GTGTCCATGCTGTGTCTGAAGAAACAGACTGGGGAACCTTATGTGACAGCATCCAC 1560
DB 1501 GTGTCCATGCTGTGTCTGAAGAAACAGACTGGGGAACCTTATGTGACAGCATCCAC 1560
QY 1561 CAGTGAACAGGATATGTGCTCTTCTTGTGATCTTCTGCTGCTGCTGCTGCTGCTG 1620
DB 1561 CAGTGAACAGGATATGTGCTCTTCTTGTGATCTTCTGCTGCTGCTGCTGCTGCTG 1620
QY 1621 AGACTTTGTGGCCTGAGGCTTATTAAGCAAGACAGATGATGAGTAATTCATTA 1680
DB 1621 AGACTTTGTGGCCTGAGGCTTATTAAGCAAGACAGATGATGAGTAATTCATTA 1680
QY 1681 ACCGCTGTCACATCTTGGCCATGAGGGAATGATCTTTCACCAAAAGAGCTCACAGC 1740
DB 1681 ACCGCTGTCACATCTTGGCCATGAGGGAATGATCTTTCACCAAAAGAGCTCACAGC 1740
QY 1741 ATTTCACAGAGATCAAAATTCAGCCCTTGGAGCTTCCAGTGTGATCAAGAGAGAA 1800
DB 1741 ATTTCACAGAGATCAAAATTCAGCCCTTGGAGCTTCCAGTGTGATCAAGAGAGAA 1800
QY 1801 GTGGGAACAAGTGTGATGCTACTATAGCTGTGACCAACAGATATCGGAATCAGA 1860
DB 1801 GTGGGAACAAGTGTGATGCTACTATAGCTGTGACCAACAGATATCGGAATCAGA 1860
QY 1861 AATATGAACAAATCTCTGCAAAAAGAGCAAGCTCTTAACTTCAAGAGTGGCTGGCC 1920

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:37:37 ; Search time 427 Seconds

(without alignments)
10716.770 Million cell updates/sec

Title: US-09-816-825-1

Sequence: 1 ggcctcgagcgccagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-----------------------------|
| 1 | 2032 | 100.0 | 2032 | 20 | AAZ20792 Human glycosyl sul |
| 2 | 1981.2 | 97.5 | 2065 | 21 | AAK94211 Human transferrase |
| 3 | 1900.8 | 93.5 | 1979 | 22 | AAK94229 Human full-length |
| 4 | 1261.6 | 62.1 | 1333 | 24 | AAK16847 Human l-selectin |
| 5 | 805.6 | 39.6 | 877 | 22 | AAK91803 Human CDNA 5'-end |
| 6 | 805.6 | 39.6 | 877 | 22 | AAK93921 Human CDNA clone r |
| 7 | 741.6 | 27.0 | 1926 | 20 | AAZ20793 Mouse glycosyl sul |
| 8 | 548.8 | 27.0 | 2988 | 21 | AAK76156 Human GREX ORF1711 |
| 9 | 506.4 | 24.9 | 517 | 24 | AAK54724 Human colon cancer |

| | | | | | |
|----|-------|------|--------|----|-----------------------------|
| 10 | 414.4 | 20.4 | 1647 | 24 | AAZ24670 Human drug metabol |
| 11 | 414.4 | 20.4 | 1694 | 22 | AAZ02700 Human glycosyl sul |
| 12 | 414.4 | 20.4 | 2544 | 24 | ABN89506 Human corneal N-ac |
| 13 | 414.4 | 20.4 | 160552 | 22 | AAZ02697 Human glycosyl sul |
| 14 | 400 | 19.7 | 2044 | 22 | AAZ02699 Human glycosyl sul |
| 15 | 400 | 19.7 | 2170 | 22 | AAZ02698 Human glycosyl sul |
| 16 | 388.8 | 19.1 | 1937 | 22 | AAK16948 Murine intestinal |
| 17 | 388.8 | 19.1 | 1989 | 22 | AAZ02696 Mouse glycosyl sul |
| 18 | 378.2 | 18.6 | 548 | 22 | AAK92588 Human CDNA 3'-end |
| 19 | 337.8 | 16.6 | 48436 | 24 | ABN89533 Human corneal N-ac |
| 20 | 287.4 | 14.1 | 389 | 24 | AAK54794 Human colon cancer |
| 21 | 143.2 | 7.0 | 2156 | 19 | AAV21200 Glycosaminoglycan |
| 22 | 142.6 | 7.0 | 668 | 24 | ABO21506 Oligonucleotide fo |
| 23 | 142.6 | 7.0 | 668 | 24 | ABO21507 Oligonucleotide fo |
| 24 | 133.6 | 6.6 | 1458 | 19 | AAV36418 Keratan sulphate 6 |
| 25 | 130 | 6.4 | 2354 | 18 | AAK45037 Chick chondroitin |
| 26 | 123.8 | 5.9 | 3029 | 22 | AAH17922 Human CDNA sequenc |
| 27 | 119.4 | 5.9 | 2393 | 24 | AAK61699 Lung small cell ca |
| 28 | 119.4 | 5.9 | 2409 | 20 | AAK87821 Human N-acetylgluc |
| 29 | 109.4 | 5.4 | 524 | 24 | ABO21510 Oligonucleotide fo |
| 30 | 109.4 | 5.4 | 524 | 24 | ABO21511 Oligonucleotide fo |
| 31 | 106.4 | 5.2 | 524 | 24 | ABO21512 Oligonucleotide fo |
| 32 | 106.4 | 5.2 | 524 | 24 | ABO21513 Oligonucleotide fo |
| 33 | 106.4 | 5.2 | 2150 | 20 | AAK87820 Mouse N-acetylgluc |
| 34 | 106 | 5.2 | 668 | 24 | ABO21508 Oligonucleotide fo |
| 35 | 106 | 5.2 | 668 | 24 | ABO21509 Oligonucleotide fo |
| 36 | 89.4 | 4.4 | 576 | 24 | ABO45408 Oligonucleotide fo |
| 37 | 89.4 | 4.4 | 576 | 24 | ABO45409 Oligonucleotide fo |
| 38 | 69.6 | 3.4 | 576 | 24 | ABO45406 Oligonucleotide fo |
| 39 | 69.6 | 3.4 | 576 | 24 | ABO45407 Oligonucleotide fo |
| 40 | 62.6 | 3.1 | 731 | 20 | AAZ24563 Human lung tumor a |
| 41 | 62.6 | 3.1 | 731 | 21 | AAZ24563 Human lung tumor a |
| 42 | 62.6 | 3.1 | 731 | 24 | ABL48021 Human lung tumor |
| 43 | 60 | 3.0 | 60 | 24 | ABN8074 Human spliced tran |
| 44 | 59.2 | 2.9 | 303 | 24 | ABN8074 Human small cell ca |
| 45 | 59.2 | 2.9 | 303 | 24 | AAK51848 Lung small cell ca |

ALIGNMENTS

RESULT 1
AAZ20792
ID AAZ20792 standard; DNA: 2032 BP.
XX AAZ20792;
AC AAZ20792;
XX 08-DEC-1999 (first entry)
DT
XX
XX Human glycosyl sulfotransferase-3 coding sequence.
DE
XX
XX Human glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW glycosyl binding interaction; inflammation; lymphocyte homing; human;
KW secondary lymph organ; ss.
XX
XX Homo sapiens.
OS
XX
XX WO9949018-A1.
PN
XX
XX 30-SEP-1999.
PD
XX
XX 26-FEB-1999; 99WO-US04316.
PF
XX
XX 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
PR
XX
XX (REGC) UNIV CALIFORNIA.
PA (SYNT) SYNTX USA INC.
XX
XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
PI WPI: 1999-580442/49.
XX
XX P-PSDB: AAK39918.
DR

```

Db 108326 CGGAGCCGCTGGCAGAAATCCGTGCGCTCTACGCTTCACTGAGGCTCAGTCTCACGCCA 108385
QY 1082 CATCTCAGACCTGGATGATTAACATCACCCGAGGCAAGGCGATGGGTG-----ACCA 1135
Db 108386 CAGCTCAGGCGCTGGATCCATACATCACCCAGGATCTGACCTGGTGGCCCGCGGAA 108445
QY 1136 GCCTTCACACAAATGCCAGGATGCCCTTAATGCTCTCCAGGCTTGGCGCTGCTTTG 1195
Db 108446 GCCTTCAGACTTCGTCCAGGAATGCCCTCAAGTCTCCAGGCTTGGCGCATGGCGTG 108505
QY 1196 CCCTATGAAAGGTTCTCGACCTTCAGAAAGCCTGTGGCGATGCGATGAATTTGCTGGGC 1255
Db 108506 CCCTTTCAGAGATCCCGCGCGTCCAGAACTGTGGCGCTGTGGCTGCACTGCTGGGC 108565
QY 1256 TACCGCCAGCTCAGATCTGAACAAGAACAGAAACCTGTGTGATCTTCTG 1309
Db 108566 TACCGCGCTGTGTACTCTGAGAGAGACGACGCAACCTCCCTTGATCTGGTG 108619

```

Search completed: January 10, 2003, 21:02:19
 Job time : 6578 secs

| QY | 1022 | CGAGCCCTGTGGCCACACACTCCCGANATGATGATTCGGGGATTCGAAATTCCTGGCC | 1081 |
|------------|--|--|----------------------------|
| Db | 70368 | CGGGGCGCGGTGGCGAAGATCCGTGCGCTTCTACGCTTCTACTGAGCTAGTCTACGCCA | 70427 |
| QY | 1082 | CATCTTCAGACCTGGGTGCATACATCAACCCGAGGCAAGGCGCATGGTG- ----ACGAC | 1135 |
| Db | 70428 | CAGCTCGAGGCGCTGGATCCATATCATCAACCCGAGGATCGACCTGTGGCGCGCGCA | 70487 |
| QY | 1136 | GCTTTCACACAAATGCCAGGATGCCCTTATATGTCCTCCAGCTTGGCCGCTGGCTTTG | 1195 |
| Db | 70488 | GCTTTCAGACCTGCTCCAGGATGCGTCAACGCTCCACGCTGGCGCATGCGCTG | 70547 |
| QY | 1196 | CCCTATGAAAAGTTCTTCGACTTCAGAAAGCCTGTGGCGATCCATGAATTCGTGGGC | 1255 |
| Db | 70548 | CCCTTGGCAGATGCGGCGCGCTGCAGGAACTGTGCGCTGCGCTGAGTGGTGCGC | 70607 |
| QY | 1256 | TACGCCCACTGACATCTGCACAAACAGAGAAACCTGTCTGATCTTGTG | 1309 |
| Db | 70608 | TACCGGCGCTGTGTACTCTGAGAGACGAGCAGCAAGCCTGCGCTTGTGCTGTG | 70661 |
| RESULT | 14 | | |
| AC009163/c | | | |
| LOCUS | AC009163 | 157337 bp | DNA linear PRI 02-NOV-2001 |
| DEFINITION | Homo sapiens chromosome 16 clone RP11-77K12, complete sequence. | | |
| ACCESSION | AC009163 | | |
| VERSION | AC009163.5 | GI:16596526 | |
| KEYWORDS | HTG. | | |
| SOURCE | Homo sapiens. | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | |
| REFERENCE | 1 (bases 1 to 157337) | | |
| AUTHORS | DOE Joint Genome Institute. | | |
| TITLE | Sequencing of Human Chromosome 16 | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 157337) | | |
| AUTHORS | DOE Joint Genome Institute. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA | | |
| REFERENCE | 3 (bases 1 to 157337) | | |
| AUTHORS | DOE Joint Genome Institute. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA | | |
| COMMENT | On Nov 2, 2001 this sequence version replaced gi:7689576. | | |
| | Sequence Quality Assessment: | | |
| | This entry has been annotated with sequence quality | | |
| | estimates computed by the Phrap assembly program. | | |
| | All manually edited bases have been reduced to quality zero. | | |
| | Quality levels above 40 are expected to have less than | | |
| | 1 error in 10,000 bp. | | |
| | Base-by-base quality values are not generally visible from the | | |
| | GenBank flat file format but are available as part | | |
| | of this entry's ASN.1 file. | | |
| | ----- | | |
| | Sequence Quality Assessment: | | |
| | This entry has been annotated with sequence quality | | |
| | estimates computed by the Phrap assembly program. | | |
| | All manually edited bases have been reduced to quality zero. | | |
| | Quality levels above 40 are expected to have less than | | |
| | 1 error in 10,000 bp. | | |
| | Base-by-base quality values are not generally visible from the | | |
| | GenBank flat file format but are available as part | | |
| | of this entry's ASN.1 file. | | |
| | ----- | | |
| FEATURES | | | |
| Source | Location/Qualifiers | | |
| | 1..157337 | | |
| | /organism="Homo sapiens" | | |
| | /db_xref="taxon:9606" | | |
| | /chromosome="16" | | |
| | /clone="RP11-77K12" | | |

[illegible]

```

Repeat_region
/rpt_family="Alu"
/rpt_type-dispersed
complement(15282..15589)
/rpt_family="Alu"
Repeat_region
/rpt_type-dispersed
15668..15718
/rpt_family="Alu"
/rpt_type-dispersed
15977..16098
/rpt_family="MIR"
Repeat_region
/rpt_type-dispersed
complement(17258..17433)
/rpt_family="MIR"
Repeat_region
/rpt_type-dispersed
complement(18001..18310)
/rpt_family="Alu"
Repeat_region
/rpt_type-dispersed
complement(18420..18591)
/rpt_family="MIR"
Repeat_region
/rpt_type-dispersed
18732..18767
/rpt_family="Simple_repeat"
/rpt_type-tandem
18967..20202
/gene="CHST5"
CDS
/note="1-GlcNAc6ST; carbohydrate sulfotransferase 5;
expressed in the small intestine and colon"
/codon_start=1
/product="intestinal
N-acetylglucosamine-6-O-sulfotransferase"
/protein_id="FACG26326.1"
/db_xref="GI:11023148"
/translation="WGMRAVPKVAHSTRPPARMMLPRSSKTVTVLLAQTTCLL
LEIIRPSPSPAGGDERVHVLVSSWRSQSSFLGQLESPHPDVFYLMERPWWITLIL
SQGSATIRLHMAVRDLMRIFLECDMPDPAVMPQSRNLAEFNNATSRALCPHMAW
PRTGISKDDVCKTCTROPFSLARACRSYSHVYLKEVRFNLQVILSDPLNLR
IVLHTRDRAVLRSEAGAPTLARNDNGIVIGTNGKWADPDLRLREVCVSHRIE
AAUICKPRTISGRVLYRFEDLAREPLAEINILALAFGLTITLPLEAHINHTIGSGSI
GKPIAEFTISRNARVNSQAWRHAFPKITURVEVCAGALQLLGIYRIVYSADQDRDL
TLDLVLPGRPHFVNSAPD"
Repeat_region
20504..20806
/rpt_family="Alu"
/rpt_type-dispersed
20869..20961
Repeat_region
/rpt_family="Al"
/rpt_type-dispersed
20962..21200
Repeat_region
/rpt_family="MER4-group"
/rpt_type-dispersed
21335..21429
Repeat_region
/rpt_family="MER4-group"
/rpt_type-dispersed
21430..21722
Repeat_region
/rpt_family="Alu"
/rpt_type-dispersed
21725..22007
Repeat_region
/rpt_family="Alu"
/rpt_type-dispersed
22008..22040
Repeat_region
/rpt_family="Simple_repeat"
/rpt_type-tandem
22041..22313
Repeat_region
/rpt_family="MER4-group"
/rpt_type-dispersed
22325..22593
Repeat_region
/rpt_family="Alu"
/rpt_type-dispersed
22594..22949
Repeat_region
/rpt_family="MER4-group"
/rpt_type-dispersed
23116..23321
Repeat_region
/rpt_family="Simple_repeat"
/rpt_type-tandem

```

```

repeat_region      23344..23400
                    /rpt_family="MaLR"
                    /rpt_type=dispersed
repeat_region      23433..23456
                    /rpt_family="Sample_repeat"
                    /rpt_type=tandem
                    complement(23457..23720)
                    /rpt_family="Alu"
                    /rpt_type=dispersed
repeat_region      complement(23789..24140)
                    /rpt_family="MER21_group"
                    /rpt_type=dispersed
                    24141..24287
repeat_region      /rpt_family="MaLR"
                    /rpt_type=dispersed
                    complement(24308..24420)
                    /rpt_family="MaLR"
                    /rpt_type=dispersed
                    complement(24467..24528)
repeat_region      /rpt_family="MER1_type"

```

[illegible]

QY 782 GTGACCGTGGTCCGGAGACCCCGGGCCGTGTCGTTCCGAGACGCCACAAGGGAGAT 841
|||||
Db 780 GTGACCTGGGGCCGACCCCGGGCCGTGTCGTTCCGAGAGCCAGACGCT 839
|||||
QY 842 CTCATGATGACAGTCCGATTTGATGAGGGCAGAGCATGAGCAAGATCAAGAAGAGAC 901
|||
Db 840 CTGGCGGCTGACAGACGATGCTGTGGCACCAGACGGCAGTGGGTGGAGGGCCGACCCC 899
|||
QY 902 CAACCTACTATGTGATGACAGGTATCTGCGCAAGACAGCTGGAGATCTCAAGACATC 961
|||
Db 900 GGCTGCGCGTGTGCGCGAGAGTGTCCGAGCCATGCGCATGCGGAGCGCCGACA 959
|||
QY 962 CAGTCTGTGCCCAAGGCCCTGACAGAACGCTACCTGTTGCGCTATGAGACCTGCT 1021
|||
Db 960 CTCAGAGCCGACCCCTTCTGTCGCGCGCTACCGGCTGTGGCTTCGAGGACCTGGCG 1019
|||
QY 1022 CGAGCCCTGTGGCCGACACTTCCGAAATGATGAAATTCGTGGGATTGGAATTCGCC 1081
|||
Db 1020 CGGAGCGCGTGGCAGAAATCCGTGCGCTACGCTTCACTGAGGCTCAGTCTACGCCA 1079
|||
QY 1082 CAGTCTGAGACCTGGGATACATACATACCCGAGGCAAGGGCATGGGTG-----ACCCAC 1135
|||
Db 1080 CAGCTCGAGGCTGATGATCCATACATACCCAGCATCTGAGCTGTGCGCGCCGGA 1139
|||
QY 1136 GCTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGCGTGTCTTTG 1195
|||
Db 1140 GCCTTCAAGATGTGTCAGGAATCGCTCAACGCTCCAGGCTGGCGCATGCGCTG 1199
|||
QY 1196 CCGTATGAAAGCTTCTCGACTTCAGAAAGCTTGGCGCATGCGCATGATTCCTGGGC 1255
|||
Db 1200 CCGTTGCAAGATCCGCGCGGTGACAGAACTGTGCGTGGTGGCTGAGCTGGCG 1259
|||
QY 1256 TACCGCACGCTCAGATCTGACAAAGAACAGAAACCTGTGCTGATCTCTTC 1309
|||
Db 1260 TACCGGCTGTGACTCTGAGAGCAGAGAGCGCAACCTGCCCTTATCTGGTG 1313
|||

RESULT 13

AF219991

LOCUS

AF219991 71503 bp DNA linear PRI 26-OCT-2000

DEFINITION

Homo sapiens intestinal N-acetylglucosamine-6-O-sulfotransferase (CHST5) and corneal N-acetylglucosamine-6-O-sulfotransferase (CHST6) genes, complete cds.

ACCESSION

AF219991

AF219991.1

GI:11023147

KEYWORDS

SOURCE

ORGANISM

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 71503)

Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T., Nakamura,T., Dotsu,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.

Mucular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene

Nat. Genet. 26 (2), 237-241 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

2 (bases 1 to 71503)

Akama,T.O. and Fukuda,M.N.

Direct Submission

Submitted (29-DEC-1999) Glycobiology Program, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES

SOURCE

1..71503

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q22"

/clone="CITB-483K2"

358..461

/rpt_family="L2"

repeat_region

358..461

/rpt_family="L2"

repeat_region /rpt_type=dispersed
complement(462..771)
/rpt_family="Alu"
/rpt_type=dispersed
772..945
repeat_region /rpt_family="L2"
/rpt_type=dispersed
1169..1338
repeat_region /rpt_family="MR"
/rpt_type=dispersed
complement(1553..1804)
/rpt_family="Alu"
/rpt_type=dispersed
2616..2916
repeat_region /rpt_family="Alu"
/rpt_type=dispersed
3509..3532
/note="similar to EST H12297 (EST cluster Hs.31147)"
3865..4028
/note="similar to EST H12297 (EST cluster Hs.31147)"
4099..4406
/rpt_family="Alu"
/rpt_type=dispersed
4793..5250
/rpt_family="L2"
/rpt_type=dispersed
complement(5251..5329)
/rpt_family="MR"
/rpt_type=dispersed
complement(5768..6080)
/rpt_family="Alu"
/rpt_type=dispersed
complement(6081..6265)
/rpt_family="Alu"
/rpt_type=dispersed
6111..6567
/note="similar to EST AA001593 (EST cluster Hs.6853)"
8035..8350
/rpt_family="Alu"
/rpt_type=dispersed
10545..11049
/note="similar to EST A1950923 (EST cluster Hs.6853)"
11043..11195
/rpt_family="Alu"
/rpt_type=dispersed
11200..11222
/rpt_family="Simple_repeat"
/rpt_type=tandem
11570..11871
/rpt_family="Alu"
/rpt_type=dispersed
complement(12065..12198)
/rpt_family="L1"
/rpt_type=dispersed
complement(12205..12493)
/rpt_family="Alu"
/rpt_type=dispersed
complement(12813..13110)
/rpt_family="Alu"
/rpt_type=dispersed
14185..20821
/gene="CHST5"
/gene="CHST5"
1014185..14228,16169..16263,17712..20821)
/gene="CHST5"
/product="intestinal N-acetylglucosamine
6-O-sulfotransferase"
14681..14713
/rpt_family="Simple_repeat"
/rpt_type=tandem
complement(14715..14990)
/rpt_family="Alu"
/rpt_type=dispersed
complement(14995..15288)
repeat_region

| | | | |
|------------|------|---|--------|
| Db | 1104 | TTTCCCGGAGGCGCCATGACGACGAGCGCGGTGTGCAACCACTGTCGGCGGCGGACGTCC | 1163 |
| Qy | 662 | TTTGAAGTGTGTGAGAAAGGCGTCCGGCTCTCTACAGCCACAGTGTGTCTCAAGAGGTGCC | 721 |
| Db | 1164 | TTACACCTGTGGCCCGGAGGCGTCCGCTCTCAACAGCAGTGTGTCTCAAGAGGTGCC | 1223 |
| Qy | 722 | TTCTTCAACCTGACATCCCTTACCCGCGTGTGAAGAAGCCCTCCCTCAACCTGATATC | 781 |
| Db | 1224 | TTCTTCAACCTGACATGTGTCTACCCGCTGTCAAGAGACCCCGGCTCAACCTACGCATC | 1283 |
| Qy | 782 | GTGCACCTGTGTCCGGAGACCCCGGCGGTTCCTTCCTCCGAAAGCACAAGAGGAGAT | 841 |
| Db | 1284 | GTGCACCTGTGTCCGGAGACCCCGGCGGTTCCTTCCTCCGAGACACCAAGGAT | 1343 |
| Qy | 842 | CTCATGATGTGACAGTCCATTTGTATGTGGGCGAGCATGAACACAATCTCAAGAGGAGC | 901 |
| Db | 1344 | CTGGGCGGTGTGAAGAGCATCTGCTGGCCACCAACGCGAGCTGGTGGAGGCCACCC | 1403 |
| Qy | 902 | CAACCTACTATGTATGATGATCATCTGTGCCAAAGCCAGCTGAGATCTTCAAGACATC | 961 |
| Db | 1404 | GACCTGTGCGGTGTGTGCGGAGGTGTGCGGTAGCCATGATGATGATGCGGAGCGCCACA | 1463 |
| Qy | 962 | CAGTCTGTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1021 |
| Db | 1464 | CTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1523 |
| Qy | 1022 | CGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1081 |
| Db | 1524 | CGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1583 |
| Qy | 1082 | CATCTTCAGACTGTGTGTATACATCACCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG | 1135 |
| Db | 1584 | CAGCTGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1643 |
| Qy | 1136 | GCTTTCACACAAAGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1195 |
| Db | 1644 | GCTTTCACACAGT | 1703 |
| Qy | 1196 | CCCTTTCGAAAGGTTTTCGACTTGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1255 |
| Db | 1704 | CCCTTTCGAAAGGTTTTCGACTTGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1763 |
| Qy | 1256 | TACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1309 |
| Db | 1764 | TACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG | 1817 |
| RESULT 12 | | | |
| AF280086 | | 3786 bp | linear |
| LOCUS | | | |
| DEFINITION | | Homo sapiens N-acetylglucosamine 6-O-sulfotransferase GST-4beta | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| MEDLINE | | | |
| PUBMED | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |

| FEATURES | | | | | |
|---------------------------|--|--|--------|-------|--|
| | 3401 | Hillview Ave., Palo Alto, CA 94304, USA | | | |
| source | | Location/Qualifiers | | | |
| | 1..3786 | /organism="Homo sapiens" | | | |
| | /db_xref="taxon:9606" | | | | |
| | /chromosome="16" | | | | |
| | /map="16q23.1" | | | | |
| | 1..3786 | | | | |
| gene | | /gene="GST4beta" | | | |
| | /note="Clusters with the intestinal N-acetylglucosamine 6-O-sulfotransferase (GST4alpha) gene in GenBank Accession Numbers AF176838 and AF176839" | | | | |
| | 189..1376 | /gene="GST4beta" | | | |
| CDS | | /note="similar to N-acetylglucosamine 6-O-sulfotransferase GST-4alpha" | | | |
| | /product="N-acetylglucosamine 6-O-sulfotransferase GST-4beta" | | | | |
| | /protein_id="AAG48244.1" | | | | |
| | /db_xref="GI:12060804" | | | | |
| | /translation="MWLPRVSTAVTALLLAQFLILFIVSRPSSPAGGAEVAHYLVLSNRSGSSSVGOLFPHQHPVFILMEPAHWITLLSGSATITHNAVRLDVRVFLCDMDVDFAVLWPRNLSDFCMAVSRAISPPCAISEAVCPICAROSFTYLARECKYISGVTVKLEVEFFNLQVLYPLDSPALMLRYHLVRPVALRSRQETAKLAREADGIVSTNGWEADPGILRVRECRSPHVAENATPKPFPLRGRIYLFEDDLAREDAELRALAYFTGCSTLPQLEAMHNTHIHSGSPARDAAFTSRNALNVSOARRHALPEAFIRIPROYEIQACALQLGLGRPVYSDEQNIALDIVLPRLNGEFMWSASTASHPRN" | | | | |
| polyA_signal | | 3106..3111 | | | |
| | /gene="GST4beta" | | | | |
| | /note="alternative signal" | | | | |
| polyA_signal | | 3734..3739 | | | |
| | /gene="GST4beta" | | | | |
| BASE COUNT | 848 a | 1030 c | 1104 g | 804 t | |
| ORIGIN | | | | | |
| Query Match | | 20.4%, Score 414.4; DB 9; Length 3786; | | | |
| Best Local Similarity | | 64.5%; Pident.No. 6.4e-96; | | | |
| Matches 654; Conservative | | 0; Mismatches 351; Indels 9; Gaps 2; | | | |
| Db | 302 | GAGCGCATGACAGTCTGTCTGTCTTCCTCCGGGCCCTTGCGTTCTTTGTGGGGCAG | 361 | | |
| | | | | | |
| Db | 303 | GCGCGCATGTCTGTCTGTCTGTCTTCCTCCGGGCCCTTGCGTTCTTTGTGGGGCCTAA | 362 | | |
| | | | | | |
| Oy | 362 | CTTTTGGGACAGCACCATGATGTTTTCTAACCTGATGAGACCGCGCTGGACGTGTGATG | 421 | | |
| | | | | | |
| Db | 363 | CTCTTAACACAGACACCAGCGATCTTCTACTAAGTAGAGCCGCGGTGACAGTGTGAC | 422 | | |
| | | | | | |
| Oy | 422 | ACCTTCAGACAGAGACCGCGCTGGATGCTGTGACATGCTGTGGCGAGTCTATACAGGGCC | 481 | | |
| | | | | | |
| Db | 423 | AACCTGTGACAGGGGAGCGCGCAACGCTGTGACATGCTGTGGCGAGTCTGTGTGCGCTCC | 482 | | |
| | | | | | |
| Oy | 482 | GTCCTCTTGTCGACATGAGCGTCTTTATGCTCAATAGAACCTGTGCCCCGGAGACAG | 541 | | |
| | | | | | |
| Db | 483 | GTCCTCTGTGTCGACATGAGCGTGTGTGATGCTCATATCG---CCTTGGCGCGCAACTG | 539 | | |
| | | | | | |
| Oy | 542 | TCCAGCCTCTTTCAGTGGGAGAACAACCGCGGCCCTGTGTTTGTGACCTGCCTGTGACATC | 601 | | |
| | | | | | |
| Db | 540 | TCCGACCTCTTCCAGTGGGCGCGTGAAGCCGTGCTGTCGCCACCGCCTTGCAAGTGC | 599 | | |
| | | | | | |
| Oy | 602 | ATCCCACAGATGGAATATCCCCCGGCGCTGACGACAGCTCTCGTGTGACATCAACAGCCC | 661 | | |
| | | | | | |
| Db | 600 | TTTCCCGAGGCGCATAGCAGCGAGGCGCTGTGTCAAGCCACTGTGTGCGCGGCGATGCC | 659 | | |
| | | | | | |
| Oy | 662 | TTTGAAGTGTGGAAGAGCCTGCGCTCTCAACAGCCAAGTGTGCTCTAAGAGAGTGGC | 721 | | |
| | | | | | |
| Db | 660 | TTTCAACCTCGGCGGAGGAGCTGCGCGCTCTCAACAGCCACGTGGTGTCAAGAGAGTGGC | 719 | | |
| | | | | | |
| Oy | 722 | TTTCTTCAACCTGCGAGTCCCTCTACCCGCTGTGTGAAGAAGCCCTCTCCATCACTGTGATTC | 781 | | |
| | | | | | |
| Db | 720 | TTTCTTCAACCTGCGAGTGTCTTACCCGCTGTGTCAAGCAAGCCCGCGCTCAACCTACGATTC | 779 | | |

```

Db 175 CTCTCAACGACGACCCCGAGCTTCTACTATGAGCCCGGTGCGACGCTGAC 234
QY 422 ACCTTCAAGACGACCCCGATGATGCTGACATGAGCTGTGCGGATGTGATACG 481
Db 235 ACCCTGTGCGAGGCGGCGGACGACCTGACATGAGCTGTGCGGACGACCTGTG 294
QY 482 GTCTTCTGTGCGACATGAGGCTTTTGATGCTGACATGAGGAGTGTGCGGAC 541
Db 295 GTCTTCTGTGCGACATGAGGCTTTTGATGCTGACATGAGGAGTGTGCGGAC 351
QY 542 TCCAGGCTCTTCACTGAGGAGACAGCGGCGCTGTGCTGACGCTGACGCTGAC 601
Db 352 TCCAGGCTCTTCACTGAGGAGACAGCGGCGCTGTGCTGACGCTGACGCTGAC 411
QY 602 ATCCACAGATGAATCATCCCGGCGCTGACGAGGCTGCTGCTGACGAGTCAAC 661
Db 412 TTTCGCCGAGCGGCGCTGACGAGGCGCTGCTGACGAGGCTGCTGCTGACG 471
QY 662 TTTCGAGGAGGAGGAGGCGCTGCTGACGAGGCGCTGCTGACGAGGAGTGC 721
Db 472 TTTCGAGGAGGAGGAGGCGCTGCTGACGAGGCGCTGCTGACGAGGAGTGC 531
QY 722 TTCTTCAACGCTGACGCTGCTGACGCTGCTGACGAGGAGGAGGAGGAGT 781
Db 532 TTCTTCAACGCTGACGCTGCTGACGCTGCTGACGAGGAGGAGGAGGAGT 591
QY 782 GTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
Db 592 GTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 651
QY 842 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 901
Db 652 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 711
QY 902 CAACCTACTATGTGATGATGATGATGATGATGATGATGATGATGATGAT 961
Db 712 GCGGAGCGGCTGCGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
QY 962 CAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1021
Db 772 CTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 831
QY 1022 CGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081
Db 832 CGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
QY 1082 CATCTTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1135
Db 892 CAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 951
QY 1136 GCTTTCACACAAATGCCAGGAGTGCCTTATGCTTCTCCAGGCTTGGCGT 1195
Db 952 GCTTTCACACAAATGCCAGGAGTGCCTTATGCTTCTCCAGGCTTGGCGT 1011
QY 1196 CCGTATGAAGGCTTTCGAGCTTCAAGAACCTGTGCGATGCAATTTCTG 1255
Db 1012 CCGTATGAAGGCTTTCGAGCTTCAAGAACCTGTGCGATGCAATTTCTG 1071
QY 1256 TACCGCCAGCTGAGTGAACAGAACAGAACCTGTGCTGAGTCTTCTG 1309
Db 1072 TACCGCCAGCTGAGTGAACAGAACAGAACCTGTGCTGAGTCTTCTG 1125

```

```

RESULT 11
AF219990 2544 bp mRNA linear PRI 26-OCT-2000
LOCUS Homo sapiens corneal N-acetylglucosamine-6-O-sulfotransferase
DEFINITION (CHST6) mRNA, complete cds.
ACCESSION AF219990.1 GI:11023145
VERSION AF219990.1
KEYWORDS Homo sapiens.

```

```

ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2544)
AUTHORS Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T.,
Nakamura,T., Dots,A., Kawasaki,S., Inoue,Y., Meda,N., Yamamoto,S.,
Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
Macular corneal dystrophy type I and type II are caused by distinct
mutations in a new sulfotransferase gene
Nat. Genet. 26 (2), 237-241 (2000)
TITLE JOURNAL MEDLINE 20472330
PUBMED 11017086
REFERENCE 2 (bases 1 to 2544)
AUTHORS Akama,T.O. and Fukuda,M.N.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-1999) Glycobiology Program, The Burnham
Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source 1..2544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22"
1..2544
/gene="CHST6"
/feature="CHST6"
/feature="responsible for macular corneal dystrophy"
693..1880
/feature="CHST6"
/feature="transfers sulfate to 6-OH of GlcNAc in keratan
sulfate"
/feature="C-GlcNAc6ST"
/feature="corneal
product_start=1
product_end=1
N-acetylglucosamine-6-O-sulfotransferase"
/protein="16-AAG26325.1"
/db_xref="GI:11023146"
/translation="MPLPVSSTAYTALLAQTFLPLVSRPSPAGGEARVHVL
VLSWRSSEFVGOLFNOHVDVFLMEPMHWTLISQSAATLIMAVRDVRSVFLC
DMVDVDAVLPFRNLSDLEFQWVRSALSPSPASAFPGALISSEVPCARQSEFL
AEERSSSHVYLKEVRFNLQVPLISDPLNRIYHIVDPAVRSRQAKAL
ARDNGVSTNNTWAEAPGLREVEKRSRRIAEATLPRPTLRRYLYREDL
ARPLAEIRALFTGLSLTPOLEKRNHITHGSGPGRREKFTSSRNALNVSQMR
HALPRAKIRVVELCAGALDILGYPVSEDEORNLADLVLPRLNFTWASSASH
PRN"
BASE COUNT 460 a 799 c 733 g 552 t
ORIGIN
Query Match 20.4%; Score 414.4; DB 9; Length 2544;
Best Local Similarity 64.5%; Pfd. No. 6.2e-96;
Matches 654; Conservative 0; Mismatches 351; Indels 9; Gaps 2.
QY 302 GAGCGATGACGCTGCGTGTGCTTCTGCTGCGCTGCGCTCTTGTGAGGCGAG 361
Db 302 GAGCGATGACGCTGCGTGTGCTTCTGCTGCGCTGCGCTCTTGTGAGGCGAG 361
QY 807 GCGCGGTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 866
Db 807 GCGCGGTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 866
QY 362 CTTTGTGGAGACCGACGATGTTTCTACCTATGAGGCGCGCTGCGACGTGTGATG 421
Db 362 CTTTGTGGAGACCGACGATGTTTCTACCTATGAGGCGCGCTGCGACGTGTGATG 421
QY 867 CTCTTCAACGACGACCGCGCTTCTTACCTATGAGGCGCGCTGCGACGTGTGAGC 926
Db 867 CTCTTCAACGACGACCGCGCTTCTTACCTATGAGGCGCGCTGCGACGTGTGAGC 926
QY 422 ACCTTCAAGAGGACCGCGCTGATGCTGACATGCTGCTGCGGATGTGATACGGGCG 481
Db 422 ACCTTCAAGAGGACCGCGCTGATGCTGACATGCTGCTGCGGATGTGATACGGGCG 481
QY 927 ACCCTGTGAGGAGGCGGCGGACGCTGCAATGCTGTGGCAGCACTGTGCGCTCTC 986
Db 927 ACCCTGTGAGGAGGCGGCGGACGCTGCAATGCTGTGGCAGCACTGTGCGCTCTC 986
QY 482 GTCTTCTGTGCGACATGAGGCTTTTGATGCTGACATGAGGAGTGTGCGGACGAG 541
Db 482 GTCTTCTGTGCGACATGAGGCTTTTGATGCTGACATGAGGAGTGTGCGGACGAG 541
QY 987 GTCTTCTGTGCGACATGAGGCTTTTGATGCTGACATGAGGAGTGTGCGGACGAG 1043
Db 987 GTCTTCTGTGCGACATGAGGCTTTTGATGCTGACATGAGGAGTGTGCGGACGAG 1043
QY 542 TCCAGGCTCTTCACTGAGGAGACAGCGGCGCTGTGCTGACGCTGCTGACATC 601
Db 542 TCCAGGCTCTTCACTGAGGAGACAGCGGCGCTGTGCTGACGCTGCTGACATC 601
QY 1044 TCCGACCTTTCAGTGGGCGGTGAGCGCTGCACTGTGCTGCGACCGCGCTGCA 1103
Db 1044 TCCGACCTTTCAGTGGGCGGTGAGCGCTGCACTGTGCTGCGACCGCGCTGCA 1103
QY 602 ATCCACAGATGAATCATCCCGGCGCTGACGAGGCTGCTGCTGCTGCTGCTGCT 661
Db 602 ATCCACAGATGAATCATCCCGGCGCTGACGAGGCTGCTGCTGCTGCTGCTGCT 661

```

QY 765 CCCTCAACCTGCATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 824
 DB 985 CCCTCAACCTGCATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1044
 QY 825 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 884
 DB 1045 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1104
 QY 885 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 944
 DB 1105 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1164
 QY 945 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1004
 DB 1165 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1224
 QY 1005 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1064
 DB 1225 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1284
 QY 1065 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1124
 DB 1285 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1344
 QY 1125 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1184
 DB 1345 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1404
 QY 1185 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1244
 DB 1405 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1464
 QY 1245 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1304
 DB 1465 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1524
 QY 1305 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1364
 DB 1525 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1584
 QY 1365 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1424
 DB 1585 AACGACAAAGGAGATATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1644
 RESULT 9
 AX381256 517 bp DNA linear PAT 18-MAR-2002
 LOCUS AX381256
 DEFINITION Sequence 194 from Patent WO0212280.
 ACCESSION AX381256
 VERSION AX381256.1 GI:19576075
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Pyle, R.A., Xu, J., and Secrist, H.
 TITLE Compositions and methods for the therapy and diagnosis of colon
 JOURNAL Patent: WO 0212280-A 194 14-FEB-2002;
 FEATURES
 LOCATION/Qualifiers
 source 1..517
 BASE COUNT 121 a 130 c 124 g 142 t
 ORIGIN
 Query Match 24.9%; Score 506.4; DB 6; Length 517;
 Best Local Similarity 99.8%; Pred. No. 8,9e-120;
 Matches 507; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1352 GAAGGCTTTCGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1411
 DB 10 GAAGGCTTTCGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 69
 QY 1412 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1471
 DB 70 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 129
 QY 1472 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1531
 DB 130 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 189
 QY 1532 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1591
 DB 190 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 249
 QY 1592 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1651
 DB 250 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 309
 QY 1652 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1711
 DB 310 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 369
 QY 1712 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1771
 DB 370 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 429
 QY 1772 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1831
 DB 430 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 489
 QY 1832 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 1891
 DB 490 TGCCCTACATCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 517
 RESULT 10
 AX327330 1647 bp DNA linear PAT 07-JAN-2002
 LOCUS AX327330
 DEFINITION Sequence 15 from Patent WO0179468.
 ACCESSION AX327330
 VERSION AX327330.1 GI:18097876
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Policky, J.L., Hatfield, A., Burford, N., Ring, H.Z., Lal, P.,
 Tribouley, C.M., Yao, M.G., Yue, H., Tang, Y.T., Patterson, C., Das, D.,
 Sanjanawala, M.S., Gandhi, A.R., Reddy, R., Khan, F.A., Baughn, M.R.,
 Ramkumar, J., Griffin, J.A., and Au-Young, J.
 TITLE Drug metabolizing enzymes
 JOURNAL Patent: WO 0179468-A 15 25-OCT-2001;
 FEATURES
 LOCATION/Qualifiers
 source 1..1647
 BASE COUNT 287 a 560 c 501 g 299 t
 ORIGIN
 Query Match 20.4%; Score 414.4; DB 6; Length 1647;
 Best Local Similarity 64.5%; Pred. No. 6e-96;
 Matches 654; Conservative 0; Mismatches 351; Indels 9; Gaps 2;
 QY 302 GAGCGGCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 361
 DB 115 GAGCGGCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 174
 QY 362 GAGCGGCTGTCACCTGTCGCGGAGACCCCGGCGCTTTCGGTCCCGAG 421

| | | | |
|----|------|---|------|
| QY | 1295 | TTTGGATCTCTCTGCTCTACCGGACTGGCCCTGAGCAATCCACTAAGGGTTGAA | 1354 |
| Db | 1206 | TTTGCGATCTTTCTGCTCTACCGGACTGGCCCTGAGCAATCCACTAAGGGTTGAA | 1265 |
| QY | 1355 | GGCTTCTCTCCACCGGCTGTCAGCCCTACGTCACCTTCTTGAAATGCTTGTGACCTTGC | 1414 |
| Db | 1266 | GGCTTCTCTCCACCGGCTGTCAGCCCTACGTCACCTTCTTGAAATGCTTGTGACCTTGC | 1325 |
| QY | 1415 | CTACATCT | 1422 |
| Db | 1326 | CTACATCT | 1333 |

| | | | | |
|------------|--|------------|------|--------|
| RESULT 7 | | | | |
| AF109155 | | | | |
| LOCUS | AF109155 | 2201 bp | mRNA | linear |
| DEFINITION | mus musculus l-selectin ligand sulfotransferase mRNA, complete cds | | | |
| ACCESSION | AF109155 | | | |
| VERSION | AF109155.1 | GI:5596405 | | |

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scuriograthi; Muridae; Murinae; Mus.
1 (bases 1 to 2201)

Yen, J. C., Iizawa, D., Tanaka, T., Miyasaka, M., Lowe, J. B., and Fukuda, M.
A novel, high endothelial venule-specific sulfotransferase
expresses 6-sulfo sialyl Lewis(x), an E-selectin ligand displayed
by CD34

JOURNAL
OF
IMMUNOLOGY
157:1934-1944 (1996)

FEATURES
REFERENCE
2 (pages 1 to 2201)
Hiraoaka, N. and Fukuda, M.
AUTHORS
TITLE
Direct Submission
JOURNAL
Submitted (23-NOV-1998) Glycobiology, The Burnham Institute, 10950
North Torrey Pines Road, La Jolla, CA 92037, USA
Location/Qualifiers

```

"organism": "Mus musculus"
/db_xref="taxon:10090"
319..1485
/function="directs expression of 6-sulfo sialyl Lewis X,
an E-selectin ligand displayed by CD34"

```

```

/codon_start=1
/product="L-selectin ligand sulfotransferase"
/protein_id="AAd9379.1"
/db_xref="gi:5596406"
/translation="IMLLKGRILMFGSOVIVALLFIMNSVHRILDSREESRPVH
IYLSWSSGSSFFGLDFGQHDVDFVIMEPAMHWMTFTSTAKLHMAYRDLRSVFI
GMSFEDAYNNPGRKQSSLFQMGSRALCSAVCCFPFAHETSSEKRRKILCGQPP
DMVEACRSHSGFVVLKEVRLSDGALPLITDPSLNLHVHLVCRDAVFRSREHTT
ELVDSVSHLIGQHLETIKEDQPYAKMIKCSQDIVAQITLPEALQDGLFLRYV
DLYRDLQITRLRYKFGVGLPLPILQITWVNNVYRGMGQOHFHTNARALNVASAMW
WELPKEKYSQDQAGEAMDLYGLVRSQDQEGNLSIDLESSHILQGVPRG"

```

| | | | | |
|-----------------------|-----------------|--------------------|------------|-------------|
| Query Match: | 38.5% | Score 782.4 | DB 10 | Length 2201 |
| Best Local Similarity | 66.8% | Pred. No. 3.9e-191 | | |
| Matches 1250; | Conservative 0; | Mismatches 561; | Indels 39; | Gaps 8; |

| | | | |
|----|-----|---|-----|
| QY | 182 | ATGCTACTGCGCTTAAAAAATGAAGCGCTCCGTGTTCTTGTTGCCAGATGGCCATCTTG | 241 |
| | | | |
| | | | |
| | | | |
| DB | 319 | ATGATGGTGTGTAAGAAAGGAGCGCTGCTGATGTTCCCGGGTCCGACGTCATCGTTGTA | 378 |
| | | | |
| | | | |
| | | | |
| QY | 242 | GCTGTATCTTCACATGTAACAGCCACAACATCACTCCCTGCTATGAAGGACACAGCCC | 301 |
| | | | |
| | | | |
| | | | |
| DB | 379 | GCTCTCTTATCATCATATGCTCCCTCCACAGACACCTTCCGAGAGGAGGAGTCCAGAG- | 437 |

438 --GCCCTGCATGCTGCTGCTCTCTTCTCCGCGCGTCAGAGATCTCTTTTGTCGGACAG 495

| Db | 616 | GTCTCTCTGTGTGACATGAGAGCGTCTTTGATGCTCCATCAATGAAACCCAGGCCCCCGGAAACAG | 675 |
|----|-----|--|-----|
| Qy | 542 | TCCAGCCCTCTTTCAGTGGGGAACACGCGGGCCCTGTGTTCGTGCACCTGCGCTGTGACATC | 601 |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| Db | 676 | TCCAGCCCTCTTCCAGTGGGGAACACGCGGGCCCTGTGTCTCAGCGCCTGTGTGTGACTTC | 735 |

| Db | 736 | TTCTCTGCCACGAGATCAGTGCTACCCACAGAGACGTGCTGGGTACACAGGCC | 795 |
|----|-----|--|-----|
| QY | 662 | TTTACGTGTGTGAGAAAGCCTGCGCTCTACAGCCACGTTGTTCAAGAGAGTGGC | 721 |

[illegible]

| DB | 916 | GTGACCCCTGGAGCCGAGACCCCCCGGGCCGTTGTCGATCCCGGGGAGCACCACCATCATGAA | 975 |
|-----|--|---|------|
| 842 | CTCATGATTACAGTCCCATTTGATGGGAGCATGACGAAACATCAAGAAAGAGAC | 901 | |
| QY | | | |
| Db | 976 | CTGTGTTGACATCATATTGTCTGAGGAGCATTTTGGAAACGATTCAGAGAGAAAGAC | 1035 |

0Y CAGTCTTGGCCCAAGGCGCTGACAGAAAGCGTACCTGCTTGAGCCCTATGAGAGACCTGGCT 1021
 1036 CAGGCCCTATTATGCGACGACGATCATTTGGCAAAAGCCAGGTGGGCATATGCTCAAGGCCATC 1095
 0Y CAGCCCTACCTATGATGATGACAGGGTATATCTCCAAAGCCAGCTGAGATCTATCAAAATCATC 501

| QY | 1022 | CGAGCCCTTGGCCAGACTTCCCGATGTATGAATTCTGGATTGGAATTTCTTGCC | 1081 |
|----|------|---|------|
| | | | |
| Db | 1156 | CGGGCACCCCTGGCCGCCAGACAGACTATTAATTTGGGGTTGGATTTTTCGCC | 1215 |
| QY | 1082 | CATCTTAGACCTGGGCGCATACATACCCGAGCAAGGCAATGGGTGACACGCTTC | 1141 |
| | | | |
| Db | 1216 | CACCTCCAAACATGGGTTCAATGTACCCGGCGCAAGGGCATAGGGTCAGACGCTTC | 1275 |
| QY | 1142 | CACCAAAATCCAGGGGATGGCCTTAATGTCTCCAGAGCTGGGCGCTGTCTTGGCCCTAT | 1201 |
| | | | |
| Db | 1276 | CATACCTAGCCCAAGAAACGCCCTCAGCGTCTTCAGAGCGTGGCGTGTGCTTACCTTAC | 1335 |

| | | | |
|----------------|------|--|------|
| D _b | 1336 | GAAAGGTTTCCAGCTTCAAAGATGCCTGGGGTGAGGCTATTGATTTGGCTGGGAATACCTC | 1395 |
| Q _y | 1262 | CACGTGAGATCTGAAACAAGAACAAGAAACCTGTGCTGATCTCTGTACTCGGACT | 1321 |
| D _b | 1396 | CAGGTGAGATCTCAACAAGAACAAGCAACCTGCCCTGGATCTTGTCCCTCCWCCAT | 1455 |
| Q _y | 1322 | GTCCTCGAGCAAATCCACTTAAGAGGGTTGGAGAAGGCTTTGCTG--CCACTGGTGTAGC | 1379 |
| D _b | 1456 | ATCTTGGGCGAGGCTCTTCCAGGAAGGTTAAGAGGGTGTGTGTGCAACCCTCTGGTTCACG | 1515 |
| | 1380 | CTTAGACACCTTTCTCTCAAAAGCCCTTCTAGGCTTCCCTCAATCTCTGTGACCCCTTAATCTCAT | 1439 |

Db 144138 GATCTTCCACCAAGAGCTACCCAGCATTTTCCACAGAGATGCAAAATCTTGAGCCCTGG 144197
QY 1775 AGTTGCCAGTGGATTTCAGGAGAGAGAGTGGGACAAAGTTGGATGCTTACTATGAGCTT 1834
Db 144198 AGTTCCCGAGTGGATTTCAGGAGAGAGTGGGACAAAGTTGGATGCTTACTATGAGCTT 144257
QY 1835 GACCATTCACAGCTATCGGTAATCAGAAATATGAAACAAATCTCTGCACAAAGAGCAG 1894
Db 144258 GACCATTCACAGCTATCGGTAATCAGAAATATGAAACAAATCTCTGCACAAAGAGCAG 144317
QY 1895 CTCTTAAGTTACAGAGGCTGGCTGGCTGCTGCTATTTGAATATCCTTCCCTGCTTTTCC 1954
Db 144318 CTCTTAAGTTACAGAGGCTGGCTGGCTGCTGCTATTTGAATATCCTTCCCTGCTTTTCC 144377
QY 1955 CATCACATAGAGAGCTTGGAGCTGAGCTGCATCTGTAAATAGTAATAATTTCCCAAT 2014
Db 144378 CATCACATAGAGAGCTTGGAGCTGAGCTGCATCTGTAAATAGTAATAATTTCCCAAT 144437
QY 2015 AAG 2017
Db 144438 AAG 144440
RESULT 6
AF149783 1333 bp mRNA linear PRI 02-JUL-2001
LOCUS AF149783
DEFINITION Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
ACCESSION AF149783
VERSION AF149783.1 GI:13897503
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1333)
AUTHORS Yeh,J.C., Hiraoka,N., Petrujak,B., Nakayama,J., Ellies,L.G.,
Rabuka,D., Hindsgraul,O., March,J.D., Lowe,J.B. and Fukuda,M.
TITLE Novel sulfated lymphocyte homing receptors and their control by a
core1 extension beta1,3-N-acetylglucosaminyltransferase
JOURNAL Cell 105 (7), 957-969 (2001)
MEDLINE 21332592
PUBMED 11439191
REFERENCE 2 (bases 1 to 1333)
AUTHORS Hiraoka,N. and Fukuda,M.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
1. 1333
/organism="Homo sapiens"
/db_xref="taxon:9606"
93..1253
/note="LST"
/codon_start=1
/product="L-selectin ligand sulfotransferase"
/protein_id="AAK48417.1"
/db_xref="GI:13897504"
/translation="MLPKMKLLFLFSOMATIALFEHMYSHNISLSKAPERMH
VILSSMRGSSFFVGLFEGHPDYFVMEAMVMTFKNSTAMMIMHAYRDLRAVF
LCKMSYDAYMEGRPRKQSSLFQWENSRALCSRPACDITFODITFRAHRLCSQP
FEVEERKARSTYSHVLEVEFENLQSLYPLKPSLNLHVLVROPRAVFRERETK
GDLMDIRIVYSGHEOKLREKEDQPYVMQICQSOLEIKYKTIGSLERADLERLVRY
EDLARAVQTSRYMEFVGLFLPHLOTVHNITRGMDHAFHINARALNVSOAM
RWSLPYKRVRLIKACGDAANLLGYRHVRESEQRNLLDLSTWVPEQIH"
BASE COUNT 286 a 393 c 336 g 318 t
ORIGIN
Query Match 62.18: Score 1261.6; DB 9; Length 1333;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1264; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 155 TCTTGCAAGCTTCGACTTCAGCACAATGCTACTGCTAAAAAAGAGAGCTCTGCTG 214
||| |||||

Db 66 TCCTTAAGGCTTTCACACTTCACACATGCTACGCTTAAAAAAATGAAAGCTCTGCTG 125
QY 215 TTTCTGTTTCCAGATGGCATCTTGCTCTATTTCTTCACATGTACAGCCACAATC 274
Db 126 TTTCTGTTTCCAGATGGCATCTTGCTCTATTTCTTCACATGTACAGCCACAATC 185
QY 275 AGCTCCCTGCTATGAAAGGACACAGCCGAGCGATGACGTGCTGCTTCTGCTCTG 334
Db 186 AGCTCCCTGCTATGAAAGGACACAGCCGAGCGATGACGTGCTGCTTCTGCTCTG 245
QY 335 CGCTCTGCTCTTCTTTTGTGGGGACGCTTTTGGGAGACCCAGATGTTTCTACCTG 394
Db 246 CGCTCTGCTCTTCTTTTGTGGGGACGCTTTTGGGAGACCCAGATGTTTCTACCTG 305
QY 395 ATGAGAGCCGCTGGACAGCTGTGATGACCTTAAACAGAGCACCCGCTGGATGCTGCAC 454
Db 306 ATGAGAGCCGCTGGACAGCTGTGATGACCTTAAACAGAGCACCCGCTGGATGCTGCAC 365
QY 455 ATGAGCTGTGCGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCG 514
Db 366 ATGAGCTGTGCGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCG 425
QY 515 TACATGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 574
Db 426 TACATGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 485
QY 575 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 634
Db 486 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 545
QY 635 TGCAGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 694
Db 546 TGCAGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 605
QY 695 AGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 754
Db 606 AGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 665
QY 755 AAAGACCCCTCCCTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 814
Db 666 AAAGACCCCTCCCTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 725
QY 815 CGTTCCGAGAGACGACAAAGGAGATCTCATGATGACAGCTGCTGCTGCTGCTGCTG 874
Db 726 CGTTCCGAGAGACGACAAAGGAGATCTCATGATGACAGCTGCTGCTGCTGCTGCTG 785
QY 875 CATGAGCAGAACTCAAGAGAGAGACCAACCTACTATGTGATGAGCTCATCTGCCAA 934
Db 786 CATGAGCAGAACTCAAGAGAGAGACCAACCTACTATGTGATGAGCTCATCTGCCAA 845
QY 935 AGCCAGCTGAGATCTCAAGAGACATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 994
Db 846 AGCCAGCTGAGATCTCAAGAGACATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 905
QY 995 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1054
Db 906 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 965
QY 1055 GAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1114
Db 966 GAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1025
QY 1115 GGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1174
Db 1026 GGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1085
QY 1175 CAGGCTTGGCGCTGCTGCTTGGCTTATGAAAAGTTTCTGACTTGAAGAGCTCTGCG 1234
Db 1086 CAGGCTTGGCGCTGCTGCTTGGCTTATGAAAAGTTTCTGACTTGAAGAGCTCTGCG 1145
QY 1235 GATGCAATGATTTGCTGGGCTACCGGACAGCTGAGATCTGAAACAGAGAACTCTG 1294
Db 1146 GATGCAATGATTTGCTGGGCTACCGGACAGCTGAGATCTGAAACAGAGAACTCTG 1205

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 183228)
 DOE Joint Genome Institute.
 AUTHOR'S
 TITLE
 JOURNAL
 COMMENT

Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

FEATURES
 source

Location/Qualifiers
 1. 183228
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-510M2"

BASE COUNT 51901 a 42098 c 40026 g 49203 t

ORIGIN

Query Match 91.4% Score 1856.6 DB 9: Length 183228:
 Best Local Similarity 95.8% Pred. No. 0:
 Matches 1859: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 155 TCTTCAAGGCTCTTCACATTCAGCATGCTACGCTAAATAAATGAAGCTCTGCTG 214
 Db 142578 TCCCTAAAGGCTCTTCACATTCAGCATGCTACGCTAAATAAATGAAGCTCTGCTG 142637
 QY 215 TTTCTGTTTCCAGATGCGCATCTGGCTCTATTCTTCACATGATACAGCCACACATC 274
 Db 142638 TTTCTGTTTCCAGATGCGCATCTGGCTCTATTCTTCACATGATACAGCCACACATC 142697
 QY 275 AGCTCCCTGTATGAAGGACAGCCGAGGCGATGACGCTGGTTCTGCTCTGG 334
 Db 142698 AGCTCCCTGTATGAAGGACAGCCGAGGCGATGACGCTGGTTCTGCTCTGG 142757
 QY 335 CGCTCTGCTCTTTTGTGGGGAGCTTTTGGGGAGACCCAGATGTTTCTACCTG 394
 Db 142758 CGCTCTGCTCTTTTGTGGGGAGCTTTTGGGGAGACCCAGATGTTTCTACCTG 142817
 QY 395 ATGGAGCCCGGCTGGAGCTGTGATGACCTTCAAGAGAGACAGCCGCTGATGCTGAC 454
 Db 142818 ATGGAGCCCGGCTGGAGCTGTGATGACCTTCAAGAGAGACAGCCGCTGATGCTGAC 142877
 QY 455 ATGGCTGTGCGGAGCTGTGATGAGGCGCTTTCTTGTGCGACATGAGCGTTTGTATGCC 514
 Db 142878 ATGGCTGTGCGGAGCTGTGATGAGGCGCTTTCTTGTGCGACATGAGCGTTTGTATGCC 142937
 QY 515 TACATGGAACCTGTGCTCCCGGAGAGATCCAGCTCTTTCAGTGGAGAACAGCCGGGCG 574
 Db 142938 TACATGGAACCTGTGCTCCCGGAGAGATCCAGCTCTTTCAGTGGAGAACAGCCGGGCG 142997
 QY 575 CTGTCTTGTGACCTGCTGTGACATCATCCACAAAGTAAATCAATCCCGGGGCTGAC 634
 Db 142998 CTGTCTTGTGACCTGCTGTGACATCATCCACAAAGTAAATCAATCCCGGGGCTGAC 143057
 QY 635 TGCAGGCTCTGTGCTGACATCAACAGCCCTTTGAGGTGTGAGAAAGGCTGCGGCTCTAC 694

Db 143058 TGCAGGCTCTGTGCTGACATCAACAGCCCTTTGAGGTGTGAGAAAGGCTGCGGCTCTAC 143117
 QY 695 AGCCAGCTGTGCTCAAGAGAGTGCGCTTCTTCACTGACGTCCTCTACCCGCTGCTG 754
 Db 143118 AGCCAGCTGTGCTCAAGAGAGTGCGCTTCTTCACTGACGTCCTCTACCCGCTGCTG 143177
 QY 755 AAAGACCCCTCCCTCAACCTGATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 814
 Db 143178 AAAGACCCCTCCCTCAACCTGATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 143237
 QY 815 CGTCCCGAGAGACGACAAAGGAGATCTCATGATTTGATGACAGTGTGATGAGGGGAG 874
 Db 143238 CGTCCCGAGAGACGACAAAGGAGATCTCATGATTTGATGACAGTGTGATGAGGGGAG 143297
 QY 875 CATGAGAGAAATCTCAAGAGAGAGACAAACCTACTATGATGATGATGATGATGATG 934
 Db 143298 CATGAGAGAAATCTCAAGAGAGAGACAAACCTACTATGATGATGATGATGATGATG 143357
 QY 935 AGCCAGCTGTGAGATCTCAAGAGACATCCAGTCTTCCCAAGGCTCTGAGAGAGCTGAC 994
 Db 143358 AGCCAGCTGTGAGATCTCAAGAGACATCCAGTCTTCCCAAGGCTCTGAGAGAGCTGAC 143417
 QY 995 CTGCTTGTGCTATGAGACCTGCTGAGAGCCCTGCTGAGAGCTTCCGATGATAT 1054
 Db 143418 CTGCTTGTGCTATGAGACCTGCTGAGAGCCCTGCTGAGAGCTTCCGATGATAT 143477
 QY 1055 GAATTCGTGGGATTTGATCTTGGCCCATCTTCAGACCTGGGATGATATCATCCGCA 1114
 Db 143478 GAATTCGTGGGATTTGATCTTGGCCCATCTTCAGACCTGGGATGATATCATCCGCA 143537
 QY 1115 GGCAGAGGCGATGGGATGACACAGCTTTCACACAAATGCGAGGATGCTTAATGCTCC 1174
 Db 143538 GGCAGAGGCGATGGGATGACACAGCTTTCACACAAATGCGAGGATGCTTAATGCTCC 143597
 QY 1175 CAGGCTTGGGCTGCTGCTTGGCCCTTGAAGAAAGTTCTGATGAGAAAGCTGCTGAC 1234
 Db 143598 CAGGCTTGGGCTGCTGCTTGGCCCTTGAAGAAAGTTCTGATGAGAAAGCTGCTGAC 143657
 QY 1235 GATGCCATGATTTGCTGGGCTACCGCCAGCTGATGATGATGATGATGATGATGATG 1294
 Db 143658 GATGCCATGATTTGCTGGGCTACCGCCAGCTGATGATGATGATGATGATGATGATG 143717
 QY 1295 TTGCTGATCTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1354
 Db 143718 TTGCTGATCTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 143777
 QY 1355 GGCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1414
 Db 143778 GGCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 143837
 QY 1415 CTACATCTGAGCCTTAACATCATGCTGTGTGGGATCACTGATGATGATGATGATGATG 1474
 Db 143838 CTACATCTGAGCCTTAACATCATGCTGTGTGGGATCACTGATGATGATGATGATGATG 143897
 QY 1475 ACAAGTGTCAAGCAGAGAGCTTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1534
 Db 143898 ACAAGTGTCAAGCAGAGAGCTTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 143957
 QY 1535 AACCTATGTGAGCAGACATCCACAGTGAAGAAAGGGATGATGATGATGATGATGATG 1594
 Db 143958 AACCTATGTGAGCAGACATCCACAGTGAAGAAAGGGATGATGATGATGATGATGATG 144017
 QY 1595 GATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1654
 Db 144018 GATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 144077
 QY 1655 ACAATATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1714
 Db 144078 ACAATATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 144137
 QY 1715 GATCTTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGCAATTTCTGAGCCCTTGG 1774

Db 88 AAGCCCAAGCCAGAGGCTTCCACTTCAGACACATGCTACGCTAATAAATAAGAGCT 147
 QY 208 CTTGCTTTTCTGTTTCCAGATGGCCATCTTGCTCTATTCTTCACATGTACAGCCA 267
 Db 148 CTTGCTTTTCTGTTTCCAGATGGCCATCTTGCTCTATTCTTCACATGTACAGCCA 207
 QY 268 CAACATGAGTCCCTGTCTATGAAGGACAGCCAGCCAGCATGAGTGTGTTCTGTC 327
 Db 208 CAATATGAGTCCCTGTCTATGAAGGACAGCCAGCCAGCATGAGTGTGTTCTGTC 267
 QY 328 TTCTGAGGCTCTGAGCTCTCTTTTGTGGGACAGCTTTTGGGACAGCCAGATGTTT 387
 Db 268 TTCTGAGGCTCTGAGCTCTCTTTTGTGGGACAGCTTTTGGGACAGCCAGATGTTT 327
 QY 388 CTACCTATGAGAGCCCTGGCAGCTGTGATGACCTTGAAGCAGAGCAGCCCTGAT 447
 Db 328 CTACCTATGAGAGCCCTGGCAGCTGTGATGACCTTGAAGCAGAGCAGCCCTGAT 387
 QY 448 GCTGACATGCTGTGAGGATCTGATAGGCGCTCTTCTTGTGACATGAGCGTCTT 507
 Db 388 GCTGACATGCTGTGAGGATCTGATAGGCGCTCTTCTTGTGACATGAGCGTCTT 447
 QY 508 TGATGCTATGATGAGACCTGTGCTGCGGAGACAGCTGAGCTCTTTCAGTGGAGAGAG 567
 Db 448 TGATGCTATGATGAGACCTGTGCTGCGGAGACAGCTGAGCTCTTTCAGTGGAGAGAG 507
 QY 568 CCGGCGCCTGTGTTGACCTGCTGCTGTGACATCATCCCAAGATGAATATCCCGC 627
 Db 508 CCGGCGCCTGTGTTGACCTGCTGCTGTGACATCATCCCAAGATGAATATCCCGC 567
 QY 628 GGTGCTATGAGAGTCTCTGTGACATCAAGCCCTTTGAGTGTGAGAGAGCCCTGCG 687
 Db 568 GGTGCTATGAGAGTCTCTGTGACATCAAGCCCTTTGAGTGTGAGAGAGCCCTGCG 627
 QY 688 CTCTACAGCCAGCTGTGCTCAAGAGAGTGTGCTTCAACCTGAGTCCCTTACCC 747
 Db 628 CTCTACAGCCAGCTGTGCTCAAGAGAGTGTGCTTCAACCTGAGTCCCTTACCC 687
 QY 748 GCTGCTAAGAGCCCTTCCCTCAACCTGATATGCTGACCTGCTCCGGAGCCCGGGC 807
 Db 688 GCTGCTAAGAGCCCTTCCCTCAACCTGATATGCTGACCTGCTCCGGAGCCCGGGC 747
 QY 808 CGTGTTCGTTCCGAGAGAGCAGCAAGAGAGATCTCATATGACAGTGTGAT 867
 Db 748 CGTGTTCGTTCCGAGAGAGCAGCAAGAGAGATCTCATATGACAGTGTGAT 807
 QY 868 GGGGAGAGATGAGCAAACTCAAGAGAGAGAGACCAACCTTACTATGTAGAGTGCAT 927
 Db 808 GGGGAGAGATGAGCAAACTCAAGAGAGAGAGAGACCAACCTTACTATGTAGAGTGCAT 867
 QY 928 CTGCAAGAGCCAGTGTGAGATCTCAAGAGACATCCATCTGCGGAGCCCTGAGGA 987
 Db 868 CTGCAAGAGCCAGTGTGAGATCTCAAGAGACATCCATCTGCGGAGCCCTGAGGA 927
 QY 988 AGCTTACCTGCTGTGCTATGAGAGCTGTGCTGAGCCCTGTGCGGAGACTTCCG 1047
 Db 928 AGCTTACCTGCTGTGCTATGAGAGCTGTGCTGAGCCCTGTGCGGAGACTTCCG 987
 QY 1048 AATGTATGAATTCGTGGATGTGAATCTTGGCCCATCTTCAAGCTTGGGTGATATCAT 1107
 Db 988 AATGTATGAATTCGTGGATGTGAATCTTGGCCCATCTTCAAGCTTGGGTGATATCAT 1047
 QY 1108 CACCCGAGGAGGAGATGGGTGACACAGCTTCCACAAATGCCAGGATGCCCTTAA 1167
 Db 1048 CACCCGAGGAGGAGATGGGTGACACAGCTTCCACAAATGCCAGGATGCCCTTAA 1107
 QY 1168 TGTCTCCAGGCTGTGGGCTGTGCTTGGCCATGAGAAAGTCTTCTGACTTCAGAAAGC 1227
 Db 1108 TGTCTCCAGGCTGTGGGCTGTGCTTGGCCATGAGAAAGTCTTCTGACTTCAGAAAGC 1167
 QY 1228 CTGTGGGAGATCCATGATTTGCTGGGCTACCCGACAGTATGATCTGAACAAGACGAG 1287
 Db 1168 CTGTGGGAGATCCATGATTTGCTGGGCTACCCGACAGTATGATCTGAACAAGACGAG 1227

QY 1288 AACCCTGTGAGATCTTCTGCTACCGAGAGCTCCCTGAGCAATCCACTAGAGG 1347
 Db 1228 AACCCTGTGAGATCTTCTGCTACCGAGAGCTCCCTGAGCAATCCACTAGAGG 1287
 QY 1348 TTGAGAGGCTTTGCGCCAGCTGTGCTGAGCCCTGAGTCTGATGATCTTCTGA 1407
 Db 1288 TTGAGAGGCTTTGCGCCAGCTGTGCTGAGCCCTGAGTCTGATGATCTTCTGA 1347
 QY 1408 GCTTGTCTACATCTCTGAGCCCTTACTATCATGTCTGTGGGATACACTGTGATGAT 1467
 Db 1348 GCTTGTCTACATCTCTGAGCCCTTACTATCATGTCTGTGGGATACACTGTGATGAT 1407
 QY 1468 TGTGTCCACAGCTGTGCTCAAGAGAGACTTTGTGTCATGCTGTGTGAGAAACAG 1527
 Db 1408 TGTGTCCACAGCTGTGCTCAAGAGAGACTTTGTGTCATGCTGTGTGAGAAACAG 1467
 QY 1528 ACTGGGAACTTATGTAGAGCAGACATCCACAGTGAAGAGGATATGCTCTTCTG 1587
 Db 1468 ACTGGGAACTTATGTAGAGCAGACATCCACAGTGAAGAGGATATGCTCTTCTG 1527
 QY 1588 TTTTCTTATCTCTGCTGTGAGAGACTTTCAGAGCTTGTGGGCTGAGAGCTATTA 1647
 Db 1528 TTTTCTTATCTCTGCTGTGAGAGACTTTCAGAGCTTGTGGGCTGAGAGCTATTA 1587
 QY 1648 GCACGACACAGTATCAGTGAATGATCCATTAACCTCCCTGCTCCACATCTTGCCCAATG 1707
 Db 1588 GCACGACACAGTATCAGTGAATGATCCATTAACCTCCCTGCTCCACATCTTGCCCAATG 1647
 QY 1708 GGGATGATCTTCTTCCCAAGAGCTCACAGACTTTTCCAGAGATGAGCAATCTGAG 1767
 Db 1648 GGGATGATCTTCTTCCCAAGAGCTCACAGACTTTTCCAGAGATGAGCAATCTGAG 1707
 QY 1768 CCCTTGGAGTCCCATGATTCAGAGAGAGTGGAGCAAGGTTGATGCCCTACTTA 1827
 Db 1708 CCCTTGGAGTCCCATGATTCAGAGAGAGTGGAGCAAGGTTGATGCCCTACTTA 1767
 QY 1828 TGAAGTGAACATCAAGATGCTGATCAATCAAAATATGAACAAATCTGCGCAAAA 1887
 Db 1768 TGAAGTGAACATCAAGATGCTGATCAATCAAAATATGAACAAATCTGCGCAAAA 1827
 QY 1888 GAGCAAGCTCTTAAGTTCACAGAGTGTGCTGGCTGATTTGAATACCTTCCCTGTC 1947
 Db 1828 GAGCAAGCTCTTAAGTTCACAGAGTGTGCTGGCTGATTTGAATACCTTCCCTGTC 1887
 QY 1948 ATTTTCCCATCATGAGAGACTTTGACCTGTGAAGCTGCCATCTGTAACTAAAT 2007
 Db 1888 ATTTTCCCATCATGAGAGACTTTGACCTGTGAAGCTGCCATCTGTAACTAAAT 1947
 QY 2008 CCCAATAGAAAAAATAAAAAA 2032
 Db 1948 CCCAATAGAGTTCTGTAAAAA 1972

RESULT 5
 AC010547 183228 bp DNA linear PRI 29-SEP-2001
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 183228)
 2 (bases 1 to 183228)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint

| | | | |
|-----------------------|--|---|------------------------|
| Db | 1741 | GGTATTCAGAAATATGCAACCAAAATCTCGCACAAAAGGCAAGCTTAAAGTTCACAG | 1800 |
| Qy | 1911 | GTGCTGGGCTGCATTTGGAATATCACTTCCCTCGCATTTTCCATCACAAGAAGCT | 1970 |
| Db | 1801 | GTGCTGGGCTGCATTTGGAATATCACTTCCCTCGCATTTTCCATCACAAGAAGCT | 1860 |
| Qy | 1971 | TTGACTGTGAAGCTGCCATCTGTATATCTAAATATCCCAATAG | 2017 |
| Db | 1861 | TTGACTGTGAAGCTGCCATCTGTATATCTAAATATCCCAATAG | 1907 |
| RESULT 4 | | | |
| LOCUS | AF280088 | 1992 bp | linear PRI 20-FEB-2001 |
| DEFINITION | Homo sapiens L-selectin ligand sulfotransferase GST-3 mRNA, | | |
| ACCESSION | AF280088 | | |
| VERSION | AF280088.1 | GI:12060807 | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | Homo sapiens. | | |
| REFERENCE | Homo sapiens. | | |
| AUTHORS | Eukariota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (bases 1 to 1992) | | |
| TITLE | Hemmerich,S., Lee,J.T.K., Bhakta,S., Bistup,A., Ruddle,N.R. and Rosen,S.D. | | |
| JOURNAL | Chromosomal localization and genomic organization for the galactose/ N-acetylgalactosamine/N-acetylglucosamine 6-0-sulfotransferase gene family | | |
| MEDLINE | Glycobiology 11 (1), 75-87 (2001) | | |
| PUBMED | 21096027 | | |
| REFERENCE | 1181564 | | |
| AUTHORS | 2 (bases 1 to 1992) | | |
| TITLE | Hemmerich,S., Bhakta,S., Lee,J.-K., Bistup,A., Ruddle,N.R. and Rosen,S.D. | | |
| JOURNAL | Direct Submission | | |
| MEDLINE | Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA | | |
| PUBMED | Location/Qualifiers | | |
| REFERENCE | 1..1992 | | |
| AUTHORS | /organism="Homo sapiens" | | |
| TITLE | /db_xref="taxon:9606" | | |
| JOURNAL | /chromosome="16" | | |
| MEDLINE | /map="16q23.1-q23.2" | | |
| PUBMED | /cell_type="high endothelial" | | |
| REFERENCE | /tissue_type="tonsil" | | |
| AUTHORS | 1..1992 | | |
| TITLE | /gene="GST3" | | |
| JOURNAL | 122..1282 | | |
| MEDLINE | /note="similar to Homo sapiens sequence encoded by GenBank Accession Number AF131235" | | |
| PUBMED | /codon_start=1 | | |
| REFERENCE | /product="L-selectin ligand sulfotransferase GST-3" | | |
| AUTHORS | /protein_id="AAG48246.1" | | |
| TITLE | /db_xref="GI:12060808" | | |
| JOURNAL | /translation="MLPKKKKLLLTIVSQMALLAFPHYSNHSLSKMAQPERNNH VLIISWSGSSSEVQGLFGQHPDVFYLMPEAMHWMTFKOSTAWMLHVAVDLIRAVF LQWMSVDAYMEQPPROSSLFQWENSRALCSAPACDIIPQDEIIPRANHLCSQOE FEVEYACRAYSHVVLKEVRFNFOSLPIPKDPSLNLIYHLVDRPAVRSRERT GDLADSRIVAGHSEKLRKDDOPYVWAVICQSLDEYKTIOSLPALQRYLLVPR EDLARIVAOISRMVEFVGLFEFLPHLOTWNVITRGCMQDHAFHTYARALNSQAN RWSIPTEKVSRLQACGDANMLDITRHYRSEQERNLDDLTSTWTVPEQIH" | | |
| BASE COUNT | 482 a | 540 c | 489 g 481 t |
| ORIGIN | | | |
| Query Match | 91.4%; | Score 1857.8; | DB 9; Length 1992; |
| Best Local Similarity | 99.1%; | Pred. No. 0; | |
| Matches 1868; | Conservative | 0; Mismatches 17; | Indels 0; Gaps 0; |
| 148 | AAGCCGCTTGCAAGTCTTTCACCTTACGACCAATGCTACTGCTATAAAAAATGAAGCT | 207 | |

Db 1321 TGTCCCTAGCAAAATCCACTAGAGGGTTGAGAGGCTTTGCTCCACCTGGTGTGAGCC 1380
 QY 1381 TCAGTCACTTTCTCTGTAATGCTTGTGAGACCTTGCCTACATCTGTAGGCTTAACATG 1440
 Db 1381 TCAGTCACTTTCTCTGTAATGCTTGTGAGACCTTGCCTACATCTGTAGGCTTAACATG 1440
 QY 1441 TCTGTGGGTATCACACTAGTGTGAGTGTGCTCCACAGTGTCTTAAGCAGAGGACTTTT 1500
 Db 1441 TCTGTGGGTATCACACTAGTGTGAGTGTGCTCCACAGTGTCTTAAGCAGAGGACTTTT 1500
 QY 1501 GTGTCCATGCTTGTGTAGAAAACAGACTGGGACCTTGTGTGAGAGACATCCAC 1560
 Db 1501 GTGTCCATGCTTGTGTAGAAAACAGACTGGGACCTTGTGTGAGAGACATCCAC 1560
 QY 1561 CAGTGAACAGGGGTATGCTCTTCTTTCTTGTATCTTCTGCTGTGGCAGACTTCA 1620
 Db 1561 CAGTGAACAGGGGTATGCTCTTCTTTCTTGTATCTTCTGCTGTGGCAGACTTCA 1620
 QY 1621 AGACTTGTGGCCTGGAGGCTTATTAAGCAGACAGTATCAGATTAATGATCATTA 1680
 Db 1621 AGACTTGTGGCCTGGAGGCTTATTAAGCAGACAGTATCAGATTAATGATCATTA 1680
 QY 1681 ACCCTCCCTGTCCACATCTTGGCCCAATGGGGAATGATCTTTCACCAAGAAGCTCACCAGC 1740
 Db 1681 ACCCTCCCTGTCCACATCTTGGCCCAATGGGGAATGATCTTTCACCAAGAAGCTCACCAGC 1740
 QY 1741 ATTTTCACAGAGATGCAATTTGAGACCTTGTGAGCTTGTGACATCAGATCGATCAGA 1800
 Db 1741 ATTTTCACAGAGATGCAATTTGAGACCTTGTGAGCTTGTGACATCAGATCGATCAGA 1800
 QY 1801 GTGGGAACAAGGTGATGAGCTTGTGAGCTTGTGACATCAGATCGATCAGA 1860
 Db 1801 GTGGGAACAAGGTGATGAGCTTGTGAGCTTGTGACATCAGATCGATCAGA 1860
 QY 1861 AATATGAACAATCTCTGTGACAAAAGAGCACTTAAAGTTCACAGGCTGCTGGGC 1920
 Db 1861 AATATGAACAATCTCTGTGACAAAAGAGCACTTAAAGTTCACAGGCTGCTGGGC 1920
 QY 1921 TGCATTTGAATATCACTTCCCTGTGATTTCCCATCAGATGAAGACTTGTGACCTG 1980
 Db 1921 TGCATTTGAATATCACTTCCCTGTGATTTCCCATCAGATGAAGACTTGTGACCTG 1980
 QY 1981 AAGCTGCATCTGTTAATCTAAATTCACCAATAGAAAAA 2032
 Db 1981 AAGCTGCATCTGTTAATCTAAATTCACCAATAGAAAAA 2032

RESULT 3
 AK026635
 LOCUS AK026635 2011 bp mRNA linear PRI 29-SEP-2000
 DEFINITION Homo sapiens cDNA: FLJ22982 fis, clone KAT11454, highly similar to AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA.
 ACCESSION AK026635
 VERSION AK026635.1 GI:10439531
 KEYWORDS oligo capping, fis (full insert sequence).
 SOURCE Homo sapiens signet-ring cell carcinoma cell_line:KATO III cDNA to mRNA, clone_lib:KAT clone:KAT11454.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Ohashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
 TITLE MEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2011)
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Ohashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAY-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human

Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: odan@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 COMMENT MEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert construction, 5'-3' end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
 FEATURES
 source Location/Qualifiers
 1..2011
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="KAT11454"
 /cell_line="KATO III"
 /cell_type="signet-ring cell carcinoma"
 /clone_lib="KAT"
 /note="cloning vector pMT185FL3"
 1..2011
 /note="highly similar to AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA"
 BASE COUNT 482 a 541 c 475 g 513 t
 ORIGIN
 Query Match 93.8%; Score 1905.4; DB 9; Length 2011;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1906; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 111 GTTACTTACACACTTCCTGGGAGAGAGTCTTTCAGACCGGCTTCAGAGGCTTCC 170
 Db 1 GTTACTTACACACTTCCTGGGAGAGAGTCTTTCAGACCGGCTTTCAGAGGCTTCC 60
 QY 171 ACTTCAGCAATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTCTTGTTTCCACA 230
 Db 61 ACTTCAGCAATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTCTTGTTTCCACA 120
 QY 231 TGGCCATCTTGGCTTATCTTCCATATGACAGCCACAATCAGTCTCTATGA 290
 Db 121 TGGCCATCTTGGCTTATCTTCCATATGACAGCCACAATCAGTCTCTATGA 180
 QY 291 AGGCACAGCCGAGAGCAGTCAAGTCTGTTCTGCTCTGGGCTGCTGCTCTT 350
 Db 181 AGGCACAGCCGAGAGCAGTCAAGTCTGTTCTGCTCTGGGCTGCTGCTCTT 240
 QY 351 TTGTGGGAGAGCTTTTGGGAGAGCAGCCAGATGTTTCTACTGATGAGCCGCTG 410
 Db 241 TTGTGGGAGAGCTTTTGGGAGAGCAGCCAGATGTTTCTACTGATGAGCCGCTG 300
 QY 411 ACCTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGCTGACATGGCTGTGGGGATC 470
 Db 301 ACCTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGCTGACATGGCTGTGGGGATC 360
 QY 471 TGATAGGGGCGCTTCTTGTGAGATGAGCGTCTTTGATGCTTACATGAACTGTG 530
 Db 361 TGATAGGGGCGCTTCTTGTGAGATGAGCGTCTTTGATGCTTACATGAACTGTG 420
 QY 531 CCGGAGAGAGTCCAGCCCTTTCAGTGGGAGAGACGCGGCGCGTGTGTGACCTG 590
 Db 421 CCGGAGAGAGTCCAGCCCTTTCAGTGGGAGAGACGCGGCGCGTGTGTGACCTG 480
 QY 591 CCTGTGACATCATCCACAGATGAATATCCCGGGCTGACTGAGGCTCTGTGCA 650
 Db 481 CCTGTGACATCATCCACAGATGAATATCCCGGGCTGACTGAGGCTCTGTGCA 540
 QY 651 GTCAAGAGCCCTTTAGGTGGTGGAGAAAGCGCTGCTTACAGCCAGCTGGTGCA 710
 Db 541 GTCAAGAGCCCTTTAGGTGGTGGAGAAAGCGCTGCTTACAGCCAGCTGGTGCA 600
 QY 711 AAGAGTGGGCTTTTCAACCTGCAAGTCCCTTACCGGCTGTGAAGACCCCTCCCA 770
 Db 601 AAGAGTGGGCTTTTCAACCTGCAAGTCCCTTACCGGCTGTGAAGACCCCTCCCA 660

FEATURES Location/Qualifiers
Source 1..2032 /organism="unknown"
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN
Query Match 100.0%; Score 2032; DB 6; Length 2032;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2032; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGTCGAGGCCAGATGCGCTCCAGCTGCGGGGAAATGCTTCCTCATTTGCTTCCAG 60
DB 1 GAGTCGAGGCCAGATGCGCTCCAGCTGCGGGGAAATGCTTCCTCATTTGCTTCCAG 60
QY 61 CCCACCTCAACAGCTCTCCCAACCCCTTGAGTCTCAGACAGTGTAAAGCTGTACTTTCA 120
DB 61 CCCACCTCAACAGCTCTCCCAACCCCTTGAGTCTCAGACAGTGTAAAGCTGTACTTTCA 120
QY 121 CAGCTTCCTGGAGCGAGTGTCTCTCAAGCCCTCTTGCAGAGTCTTCCACTTCAGCAC 180
DB 121 CAGCTTCCTGGAGCGAGTGTCTCTCAAGCCCTCTTGCAGAGTCTTCCACTTCAGCAC 180
QY 181 AATGCTACTGCTTAAATAATGAAGCTCTGCTGTTTCTGTTTCCAGATGCGCATCTT 240
DB 181 AATGCTACTGCTTAAATAATGAAGCTCTGCTGTTTCTGTTTCCAGATGCGCATCTT 240
QY 241 GAGCTCTATCTTCCACATGTRACAGCCACACACATCAGCTCCCTGCTCTGAGGACAGCC 300
DB 241 GAGCTCTATCTTCCACATGTRACAGCCACACACATCAGCTCCCTGCTCTGAGGACAGCC 300
QY 301 CGAGCGCATGACGCTGCTGCTTCTGTTCTTGGCGCTGCGCTCTTCTTTTGAGGACA 360
DB 301 CGAGCGCATGACGCTGCTGCTTCTGTTCTTGGCGCTGCGCTCTTCTTTTGAGGACA 360
QY 361 GCTTTTGGGAGAGACCCAGATGTTTCTTAOCGTGATGAGACCCCGCTGGACAGTGTGAT 420
DB 361 GCTTTTGGGAGAGACCCAGATGTTTCTTAOCGTGATGAGACCCCGCTGGACAGTGTGAT 420
QY 421 GACCTTCAAGCAGACACCGCCTGATGCTGCACATGGCTGTGGGAGATGTGATACGGGC 480
DB 421 GACCTTCAAGCAGACACCGCCTGATGCTGCACATGGCTGTGGGAGATGTGATACGGGC 480
QY 481 CGTCTTCTTGTGCACATGAGCCGCTTTGATGCTTACATGGAACCTGCTCCCGAGACA 540
DB 481 CGTCTTCTTGTGCACATGAGCCGCTTTGATGCTTACATGGAACCTGCTCCCGAGACA 540
QY 541 GTCCAGCCTCTTTCAGTGGAGAAACACCGGGCCCTGTGTTCTGCACCTGCTGTGACAT 600
DB 541 GTCCAGCCTCTTTCAGTGGAGAAACACCGGGCCCTGTGTTCTGCACCTGCTGTGACAT 600
QY 601 CATCCCAAGATGAAATTCATCCCGCGGCTCACGTGACGCTCTCTGTGACGTACAAGCC 660
DB 601 CATCCCAAGATGAAATTCATCCCGCGGCTCACGTGACGCTCTCTGTGACGTACAAGCC 660
QY 661 CTTTGAAGTGTGAGAGAGGCTGCCCTCTTACAGCCAGGTGTGTCACAGAGGTGCG 720
DB 661 CTTTGAAGTGTGAGAGAGGCTGCCCTCTTACAGCCAGGTGTGTCACAGAGGTGCG 720
QY 721 CTTCTTAACTGACGTCCCTCTTACCCGCTGTGAAAACCCCTCCCTCAACTGCATAT 780
DB 721 CTTCTTAACTGACGTCCCTCTTACCCGCTGTGAAAACCCCTCCCTCAACTGCATAT 780
QY 781 CGTGCACCTGTGCGGGAGCCCGGGGCGTGTCCGTTCCCGAGAGCCACAAGGAGGA 840
DB 781 CGTGCACCTGTGCGGGAGCCCGGGGCGTGTCCGTTCCCGAGAGCCACAAGGAGGA 840
QY 841 TCTCATGATTGACAGTGCATGTGATGAGGAGCATGAGCAGAAAATCTCAAGAGAGGA 900
DB 841 TCTCATGATTGACAGTGCATGTGATGAGGAGCATGAGCAGAAAATCTCAAGAGAGGA 900
QY 901 CCAACCTTACTATGTGATGAGGTTCATCTGCCAAAGCCAGCTGGAGATCTACAGACCAT 960
DB 901 CCAACCTTACTATGTGATGAGGTTCATCTGCCAAAGCCAGCTGGAGATCTACAGACCAT 960

QY 961 CCAGTCTTGGCCCAAGGCCCTGCGAGAACGCTACCTGTTTGCCGTATGAGACCTGGC 1020
DB 961 CCAGTCTTGGCCCAAGGCCCTGCGAGAACGCTACCTGTTTGCCGTATGAGACCTGGC 1020
QY 1021 TCGAGCCCTCTGTGGCCCAAGCTTCCGAATGTATGATTCGTGGGATTTGGAAATCTTGGC 1080
DB 1021 TCGAGCCCTCTGTGGCCCAAGCTTCCGAATGTATGATTCGTGGGATTTGGAAATCTTGGC 1080
QY 1081 CCAATTTCAAGCTTGGCTGATACATCAACCCGAGCAAGGGGATGGGATGACACGCTT 1140
DB 1081 CCAATTTCAAGCTTGGCTGATACATCAACCCGAGCAAGGGGATGGGATGACACGCTT 1140
QY 1141 CCACACAATGCCAGGAGATGCCCTTATGTCTTCCAGGCTTGCGCTGCTTTGGCCCTA 1200
DB 1141 CCACACAATGCCAGGAGATGCCCTTATGTCTTCCAGGCTTGCGCTGCTTTGGCCCTA 1200
QY 1201 TGAAGAGTTTCTCGACTTCAGAAAGCCTGTGGCGCATGGCATGAATTTGCTGGCTACCG 1260
DB 1201 TGAAGAGTTTCTCGACTTCAGAAAGCCTGTGGCGCATGAATTTGCTGGCTACCG 1260
QY 1261 CCACGTCAATCTGAACAAGAACAGAAACCTGTGCTGATCTTCTGCTTACCTGGAC 1320
DB 1261 CCACGTCAATCTGAACAAGAACAGAAACCTGTGCTGATCTTCTGCTTACCTGGAC 1320
QY 1321 TGTCCCTGAGCAAAATCCACTAAGAGGCTTGAGAAGGCTTGTCTGCCACCTGCTGTACGC 1380
DB 1321 TGTCCCTGAGCAAAATCCACTAAGAGGCTTGAGAAGGCTTGTCTGCCACCTGCTGTACGC 1380
QY 1381 TCAGTCACTTCTCTGAATGCTCTGAGCCTTGACATCTCTGAGCCTTAACTACATG -1440
DB 1381 TCAGTCACTTCTCTGAATGCTCTGAGCCTTGACATCTCTGAGCCTTAACTACATG -1440
QY 1441 TCTGTGGTATACACATGCTGAGTGTGTCCACACCTGCTCAAGCAGANGACTTTT 1500
DB 1441 TCTGTGGTATACACATGCTGAGTGTGTCCACACCTGCTCAAGCAGANGACTTTT 1500
QY 1501 GTGTCCATGCTTGTGCTGAGAAACAGACTGGGGAACCTTATGTGAGCAGACATCCGC 1560
DB 1501 GTGTCCATGCTTGTGCTGAGAAACAGACTGGGGAACCTTATGTGAGCAGACATCCGC 1560
QY 1561 CAGTGAACAGGGTATGCTCTTCTTCTTGTGATCTTCTGCTGTGCGGAGACTTGA 1620
DB 1561 CAGTGAACAGGGTATGCTCTTCTTCTTGTGATCTTCTGCTGTGCGGAGACTTGA 1620
QY 1621 AGACTTGTGGCTGAGGCTTATTAAGCAGACACAGATCAGTGAATTTGATCCATAA 1680
DB 1621 AGACTTGTGGCTGAGGCTTATTAAGCAGACACAGATCAGTGAATTTGATCCATAA 1680
QY 1681 ACCCTCCTGTCACATCTTGCCCAATGGGGAATGATCTTTACCAAAAGGCTCACAGC 1740
DB 1681 ACCCTCCTGTCACATCTTGCCCAATGGGGAATGATCTTTACCAAAAGGCTCACAGC 1740
QY 1741 ATTTTCCAGAGATGCAATTTCTAGCCCTTGAGGTTTCCCGTGGATTTCAAGGAAGAA 1800
DB 1741 ATTTTCCAGAGATGCAATTTCTAGCCCTTGAGGTTTCCCGTGGATTTCAAGGAAGAA 1800
QY 1801 GTGGGAACAAGGTGATGCTTACTTATGAGGCTTACCATCAGACTATCGGTAATTCAGA 1860
DB 1801 GTGGGAACAAGGTGATGCTTACTTATGAGGCTTACCATCAGACTATCGGTAATTCAGA 1860
QY 1861 AATATGAAACAANAATCTCGACAAAAGCAAGCTCTTAAGTTACAGAGGTGCTGGGC 1920
DB 1861 AATATGAAACAANAATCTCGACAAAAGCAAGCTCTTAAGTTACAGAGGTGCTGGGC 1920
QY 1921 TGCATTTGAATATCACTTCCCTCTGCATTTCCCATCAGATAGAAGCTTGAACCTGTG 1980
DB 1921 TGCATTTGAATATCACTTCCCTCTGCATTTCCCATCAGATAGAAGCTTGAACCTGTG 1980
QY 1981 AAGCTGCACTCTGTTPAINDCTAATAATCCCAATTAAGAAAAAATTTGAAAAA 2032
DB 1981 AAGCTGCACTCTGTTPAINDCTAATAATCCCAATTAAGAAAAAATTTGAAAAA 2032

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 17:51:20 ; Search time 5293 Seconds
(without alignments) 11172.662 Million cell updates/sec

Title: US-09-816-825-1

Perfect score: 1 ggcgcagagccagcagatgcct.....ataagaaaaaaaaaaaaa 2032

Sequence: 1 ggcgcagagccagcagatgcct.....ataagaaaaaaaaaaaaa 2032

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank : *

1: gb_ba : *

2: gb_htg : *

3: gb_in : *

4: gb_om : *

5: gb_ov : *

6: gb_pat : *

7: gb_ph : *

8: gb_pl : *

9: gb_pr : *

10: gb_ro : *

11: gb_sts : *

12: gb_sy : *

13: gb_un : *

14: gb_vi : *

15: em_ba : *

16: em_fun : *

17: em_hum : *

18: em_in : *

19: em_mu : *

20: em_om : *

21: em_ov : *

22: em_ov : *

23: em_pat : *

24: em_ph : *

25: em_pl : *

26: em_ro : *

27: em_sts : *

28: em_un : *

29: em_vi : *

30: em_htg_hum : *

31: em_htg_iny : *

32: em_htg_other : *

33: em_htg_mus : *

34: em_htg_pln : *

35: em_htg_rcd : *

36: em_htg_mam : *

37: em_htg_vrt : *

38: em_sy : *

39: em_htgo_hum : *

40: em_htgo_mus : *

41: em_htgo_other : *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 2032 | 100.0 | 2032 | 6 | AR203335 |
| 2 | 2032 | 100.0 | 2032 | 9 | AR131235 |
| 3 | 1905.4 | 93.8 | 2011 | 9 | AK026635 |
| 4 | 1857.8 | 91.4 | 1992 | 9 | AF280088 |
| 5 | 1856.6 | 91.4 | 18328 | 9 | AC010547 |
| 6 | 1261.6 | 62.1 | 1333 | 9 | AF149783 |
| 7 | 782.4 | 38.5 | 2201 | 10 | AF109155 |
| 8 | 741.6 | 36.5 | 1926 | 10 | AF131236 |
| 9 | 506.4 | 24.9 | 517 | 6 | AX381256 |
| 10 | 414.4 | 20.4 | 1647 | 6 | AX327330 |
| 11 | 414.4 | 20.4 | 2544 | 9 | AF219990 |
| 12 | 414.4 | 20.4 | 3786 | 9 | AF280086 |
| 13 | 414.4 | 20.4 | 17503 | 9 | AF219991 |
| 14 | 414.4 | 20.4 | 15737 | 9 | AC009163 |
| 15 | 414.4 | 20.4 | 208185 | 2 | AC009105 |
| 16 | 400 | 19.7 | 1462 | 9 | AF176839 |
| 17 | 400 | 19.7 | 2170 | 9 | AF176838 |
| 18 | 400 | 19.7 | 3278 | 9 | AF246718 |
| 19 | 400 | 19.7 | 194832 | 9 | AC025287 |
| 20 | 388.8 | 19.1 | 1740 | 10 | AF176841 |
| 21 | 388.8 | 19.1 | 1989 | 10 | AF176840 |
| 22 | 388.8 | 19.1 | 159170 | 2 | AC115914 |
| 23 | 388.8 | 19.1 | 237950 | 2 | AC127315 |
| 24 | 368.8 | 18.1 | 133902 | 2 | AC095664 |
| 25 | 324 | 15.9 | 108261 | 2 | AC130174 |
| 26 | 287.4 | 14.1 | 389 | 6 | AX381326 |
| 27 | 266 | 13.1 | 184551 | 2 | AC117869 |
| 28 | 146.4 | 7.2 | 6961 | 9 | AB017815 |
| 29 | 146.4 | 7.2 | 178253 | 9 | AC073370 |
| 30 | 146.4 | 7.2 | 196467 | 9 | AC022392 |
| 31 | 143.2 | 7.0 | 2156 | 6 | AR071396 |
| 32 | 143.2 | 7.0 | 2156 | 6 | E14937 |
| 33 | 143.2 | 7.0 | 2731 | 9 | AB012192 |
| 34 | 135.2 | 6.7 | 200292 | 2 | AC087442 |
| 35 | 133.6 | 6.6 | 1241 | 9 | AF090137 |
| 36 | 133.6 | 6.6 | 1458 | 6 | E16306 |
| 37 | 133.6 | 6.6 | 2190 | 6 | AR211733 |
| 38 | 133.6 | 6.6 | 2190 | 9 | HS075637 |
| 39 | 133.6 | 6.6 | 2415 | 9 | AB003791 |
| 40 | 133.6 | 6.6 | 2453 | 9 | BC028235 |
| 41 | 133.6 | 6.6 | 2734 | 9 | BC022567 |
| 42 | 130 | 6.4 | 2354 | 5 | CHKG65T |
| 43 | 130 | 6.4 | 2354 | 6 | AR050577 |
| 44 | 130 | 6.4 | 2354 | 6 | E12369 |
| 45 | 123.8 | 6.1 | 2227 | 9 | AF083066 |

ALIGNMENTS

RESULT 1

AR203335

LOCUS AR203335 2032 bp DNA linear PAT 20-JUN-2002

DEFINITION Sequence 2 from patent US 6365365.

ACCESSION AR203335

VERSION AR203335.1 GI:21499698

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2032)

AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.

TITLE Method of determining whether an agent modulates glycosyl sulfotransferase-3

JOURNAL Patent: US 6365365-A 2 02-APR-2002;